GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

CM protein - protein search, using sw model

Run on: October 29, 2003, 18:29:08 ; Search time 52.0213 Seconds
(without alignments)
3217.546 Million cell updates/sec

Title: US-09-980-043-2
Perfect score: 5240
Sequence: 1 MMESGAPICHTCGEQVGHDA.....PFVNKVDNILVAETCISIDC 978

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 642050 seqs, 171146064 residues

Total number of hits satisfying chosen parameters: 642050

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4524	86.3	974	10	US-09-838-539-6	Sequence 6, Appli
2	3959.5	75.6	1007	12	US-10-209-059-28	Sequence 28, Appl
3	3408	65.0	1078	12	US-10-209-059-26	Sequence 26, Appl
4	3367	64.3	1039	10	US-09-900-237-14	Sequence 14, Appl
5	3366.5	64.2	1079	12	US-10-267-459-6	Sequence 6, Appli
6	3366.5	64.2	1079	12	US-10-209-059-22	Sequence 22, Appl
7	3365.5	64.2	1077	12	US-10-209-059-10	Sequence 10, Appl
8	3365.5	64.2	1077	12	US-10-160-719-6	Sequence 6, Appli
9	3365.5	64.2	1077	12	US-10-160-719-30	Sequence 30, Appl
10	3362.5	64.2	1076	12	US-10-209-059-14	Sequence 14, Appl
11	3362.5	64.2	1076	12	US-10-160-719-58	Sequence 58, Appl
12	3352	64.0	1148	10	US-09-900-237-4	Sequence 4, Appli
13	3348.5	63.9	1091	10	US-09-900-237-26	Sequence 26, Appl
14	3347	63.9	1081	15	US-10-229-193-6	Sequence 6, Appli
15	3343	63.8	1081	15	US-10-229-193-12	Sequence 12, Appl
16	3342	63.8	1059	12	US-10-209-059-42	Sequence 42, Appl
17	3342	63.8	1059	12	US-10-160-719-2	Sequence 2, Appli
18	3338.5	63.7	1075	12	US-10-209-059-2	Sequence 2, Appli
19	3338.5	63.7	1075	12	US-10-160-719-10	Sequence 10, Appl
20	3338.5	63.7	1075	12	US-10-160-719-34	Sequence 34, Appl
21	3338.5	63.7	1075	12	US-10-160-719-54	Sequence 54, Appl
22	3332.5	63.6	1086	10	US-09-900-237-10	Sequence 10, Appl
23	3332.5	63.6	1086	12	US-10-209-059-50	Sequence 50, Appl
24	3332.5	63.6	1086	12	US-10-160-719-18	Sequence 18, Appl
25	3332.5	63.6	1086	12	US-10-160-719-38	Sequence 38, Appl
26	3332	63.6	1074	12	US-10-209-059-46	Sequence 46, Appl
27	3332	63.6	1074	12	US-10-160-719-14	Sequence 14, Appl
28	3332	63.6	1074	12	US-10-160-719-22	Sequence 22, Appl
29	3332	63.6	1074	12	US-10-160-719-42	Sequence 42, Appl
30	3332	63.6	1094	12	US-10-209-059-18	Sequence 18, Appl
31	3332	63.6	1094	12	US-10-160-719-26	Sequence 26, Appl
32	3332	63.6	1094	12	US-10-160-719-46	Sequence 46, Appl
33	3332	63.6	1165	10	US-09-900-237-8	Sequence 8, Appli
34	3329	63.5	1065	15	US-10-229-193-10	Sequence 10, Appl
35	3327	63.5	1065	10	US-09-900-237-33	Sequence 33, Appl
36	3326	63.5	1080	10	US-09-900-237-30	Sequence 30, Appl
37	3282.5	62.6	1043	12	US-10-160-719-50	Sequence 50, Appl
38	3242	61.9	1084	15	US-10-229-193-8	Sequence 8, Appli
39	3240.5	61.8	881	10	US-09-838-539-8	Sequence 8, Appli
40	3200.5	61.1	927	12	US-10-209-059-30	Sequence 30, Appl
41	3156	60.2	821	12	US-10-267-459-2	Sequence 2, Appli
42	3156	60.2	821	12	US-10-209-059-6	Sequence 6, Appli
43	3056	58.3	793	10	US-09-900-237-18	Sequence 18, Appl
44	2851.5	54.4	740	10	US-09-900-237-24	Sequence 24, Appl
45	2733	52.2	685	10	US-09-838-539-7	Sequence 7, Appli

ALIGNMENTS

RESULT 1
US-09-838-539-6
; Sequence 6, Application US/09838539
; Patent No. US20020129401A1
; GENERAL INFORMATION:
; APPLICANT: Stalker, D. et al.
; TITLE OF INVENTION: Plant Cellulose Synthase and Promoter
; TITLE OF INVENTION: Sequences
; FILE REFERENCE: 15621/03/US
; CURRENT APPLICATION NUMBER: US/09/838,539
; CURRENT FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/029,987
; PRIOR FILING DATE: 1996-10-29
; PRIOR APPLICATION NUMBER: 08/960,048
; PRIOR FILING DATE: 1997-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 974
; TYPE: PRT
; ORGANISM: Gossypim hirsutum
US-09-838-539-6

Query Match 86.3%; Score 4524; DB 10; Length 974;
Best Local Similarity 85.5%; Pred. No. 0;
Matches 840; Conservative 70; Mismatches 60; Indels 12; Gaps 7;

```
Qy          1 MMESGAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEINEGRKVCLRCGSPYDE 60
              ||||| |:|||||| || : ||| |||||||::|:||||||::: |||| ||||||||||
Db          1 MMESGVPVCHTCGEHVGLNVNGEPFVACHECNFPICKSCFEYDLKEGRKACLRCGSPYDE 60

Qy         61 NLLDDVEKKGSGNQSTMASHLNDSQDVGIHARHISSVSTVDSEMNDEYGNPIWKNRVKSC 120
              ||||||| | :|:|||||:||| ||||||||||||||||:|||| :: || ||||||:|
Db         61 NLLDDVE-KATGDQSTMAAHLNKSQDVGIHARHISSVSTLDSEMAEDNGNSIWKNRVESW 119

Qy        121 KDKENKKKKRSPKAETEPAQVPTEQQMEEKPSAEASEPLSIVYPIPRNKLTPYRAVIIMR 180
              |:|:||||| :       |::| |||||:||: :||:||| : |||:::| ||| |||||
Db        120 KEKKNKKKKPATTKVEREAEIPPEQQMEDKPAPDASQPLSTIIPIPKSRLAPYRTVIIMR 179
```

```
y       181 LVILGLFFHFRITNPVDSAFGLWLTSVICEIWFAFSWVLDQFPKWNPVNRETYIERLSAR 240
            |:|||||||:|:|||||||||||||||||||||||||||||||||||| |:|||||||:|:|||
b       180 LIILGLFFHYRVTNPVDSAFGLWLTSVICEIWFAFSWVLDQFPKWYPVNRETYIDRLSAR 239

y       241 YEREGEPSQLAGVDFFVSTVDPLKEPPLITANTVLSILAVDYPVDKVSCYVSDDGAAMLS 300
            ||||||| :|| |||||||||||||||||||||||||||||:||||||||||:||||||||:
b       240 YEREGEPDELAAVDFFVSTVDPLKEPPLITANTVLSILALDYPVDKVSCYISDDGAAMLT 299

y       301 FESLVETAEFARKWVPFCKKFSIEPRAPEFYFSQKIDYLKDKVQPSFVKERRAMKRDYEE 360
            ||||||||:|||||||||||||||||||||||||||||||||||||||||||||||||||||
b       300 FESLVETADFARKWVPFCKKFSIEPRAPEFYFSQKIDYLKDKVQPSFVKERRAMKRDYEE 359

y       361 YKVRVNALVAKAQKTPEEGWTMQDGTPWPGNNTRDHPGHDSGLPWEIL---GARDIEGNE 417
            ||:|:||||||||||:||||||||| ||||| |||| |:  |  |||||||||
b       360 YKIRINALVAKAQKTPDEGWTMQDGTSWPGNNPRDHP----GMIQVFLGYSGARDIEGNE 415

y       418 LPRLVYVSREKRPGYQHHKKAGAENALVRVSAVLTNAPYILNVDCDHYVNNSKAVREAMC 477
            |||||||||||||||||||||||||||||||||||||:|||:|||||||||||||||||||
b       416 LPRLVYVSREKRPGYQHHKKAGAENALVRVSAVLTNAPFILNLDCDHYVNNSKAVREAMC 475

y       478 ILMDPQVGRDVCYVQFPQRFDGIDKSDRYANRNVVFFDVNMKGLDGIQGPVYVGTGCVFN 537
            |||||||||||||||||||||||:|||||||| ||||||||||||||||||||||||||||
b       476 FLMDPQVGRDVCYVQFPQRFDGIDRSDRYANRNTVFFDVNMKGLDGIQGPVYVGTGCVFN 535

y       538 RQALYGYGPPSMPSLRKRKDSSSCFSCCCPSKKKPAQDPAEVYRDAKREDLNAAIFNLTE 597
            |||||||||||||  |   |||| ||||| ||:| :||:|:|||||||:|:|||||| |
b       536 RQALYGYGPPSMPSFPK-SSSSSC-SCCCPGKKEP-KDPSELYRDAKREELDAAIFNLRE 592

y       598 IDNYDEHERSMLISQLSFEKTFGLSSVFIESTLMENGGVPESANSPPFIKEAIQVIGCGY 657
            |||||:|||||||| |||||||||||||||||||||| ||||    ||||| || |||
b       593 IDNYDEYERSMLISQTSFEKTFGLSSVFIESTLMENGGVAESANPSTLIKEAIHVISCGY 652

y       658 EEKTEWGKQIGWIYGSVTEDILSGFKMHCRGWRSIYCMPVRPAFKGSAPINLSDRLHQVL 717
            |||| |||:|||||||||||||:|||||||||||||||:|||||||||||||||||||||||
b       653 EEKTAWGKEIGWIYGSVTEDILTGFKMHCRGWRSIYCMPLRPAFKGSAPINLSDRLHQVL 712

y       718 RWALGSVEIFFSRHCPLWYGFGGGRLKWLQRLAYINTIVYPFTSLPLIAYCTIPAVCLLT 777
            |||||||||| |||||||||||||||||||||||||||||||||||||||||::||:||||
b       713 RWALGSVEIFLSRHCPLWYGFGGGRLKWLQRLAYINTIVYPPTSLPLIAYCSLPAICLLT 772

y       778 GKFIIPTLSNLASMLFLGLFISIIVTAVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAV 837
            ||||||||||||:||||||||||||||||||||||||||||||||||||||||||||||||
b       773 GKFIIPTLSNLASVLFLGLFLSIIVTAVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAV 832

y       838 FQGFLKMLAGIDTNFTVTAKAAEDAEFGELYMVKWTTLLIPPTTLLIINMSG-CAGPSDA 896
            |||||||||||||||||||||:||:|||||:|||||||||||||||:|| |  ||||||
b       833 FQGFLKMLAGIDTNFTVTAKAADDADFGELYIVKWTTLLIPPTTLLIVNMVGVVAGFSDA 892

y       897 LNKGYEAWGPLFGKVFFAFWVILHLYPFLKGLMGRQNLTPTIVVLWSVLLASVFSLVWVK 956
            ||||||||||||||||:|||||||||||||||||||| ||||||||||||||||||||||:
b       893 LNKGYEAWGPLFGKVFFSFWVILHLYPFLKGLMGRQNRTPTIVVLWSVLLASVFSLVWVR 952

y       957 INPFVNKVDNTLVAETCISIDC 978
            |||||:  |:| |:::||||||
b       953 INPFVSTADSTTVSQSCISIDC 974
```

```
ESULT 2
JS-10-209-059-28
  Sequence 28, Application US/10209059
  Publication No. US20030163838A1
  GENERAL INFORMATION:
   APPLICANT: Dhugga, Kanwarpal S.
   APPLICANT:  Wang, Haiyin
   TITLE OF INVENTION: Maize Cellulose Synthases and Uses
   TITLE OF INVENTION:  Thereof
   FILE REFERENCE: 0864R2
   CURRENT APPLICATION NUMBER: US/10/209,059
   CURRENT FILING DATE:  2002-07-31
   PRIOR APPLICATION NUMBER: 60/096,822
   PRIOR FILING DATE: 1998-08-17
   PRIOR APPLICATION NUMBER: 09/371,383
;  PRIOR FILING DATE: 1999-08-06
;  PRIOR APPLICATION NUMBER: 08/550,483
;  PRIOR FILING DATE: 2000-04-14
;  NUMBER OF SEQ ID NOS: 52
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 28
;   LENGTH: 1007
;   TYPE: PRT
;   ORGANISM: Zea mays
US-10-209-059-28

  Query Match             [illegible]% (%);  Score 3853.5;  DB 12;  Length 1007;
  Best Local Similarity   [illegible]%;  Pred. No. 0;
  Matches  745;  Conservative   89;  Mismatches  126;  Indels   31;  Gaps   11;
```

```
Qy          1 MMES-GAPICHTCPEQVCHDANGELFVACHECGYPVCKSCFEFEINEGRKVCLRCGSPY  58
              ||:  |  |  :  :   .|      || .|| :::| :  :  ||:  | ||   |
Db          1 MMESAAAQSAAGHE    DAP    AACPACGYALCRACLDEDAAEGRTTCARCGGDYA  52

Qy         59 -----        [illegible]SMASHLNDSQDVGIHARHISSVSTVDSEM 104
                :  :::|    :  :   || :| ||   |||  :,:| :  |:
Db         53 AINPARANPYEA-PFFVFNDHTAPIPEEVDMASHLNDRQDEVSHARTMSSLSGIGSEL 112

Qy        105 NDEYGNPIWKNRVES-KDKENKKKKRSPKAEIPPAQVPTEQQ-MEEKPSAEASEPLSIVV 163
              ||| | |||:.||:  |:|:::||  : |  :  | .:| |:|'    :. '||| |
Db        113 NDESGKPIWKNRVESWKEKKKNEKKASAKKTAAKAQPPPVEEQIMDEKDLTDAYEPLSPVI 172

Qy        164 PIPRNKLTPYRAVIVMRLVILGLFFHFRITNPVDSAFGLWLTSVICEIWFAFSWVLDQFP 223
              || :|:||| :||| '|||::|||| :|||:|||||:|||||||||||||||||||||||
Db        173 PISKNKLTPYRAVIVMRLIVLGLFFHYRITNPVNSAFGLWMTSVICEIWFGFSWILDQFP 232

Qy        224 KWNPVNRETYIERLSARYEREGEPSQLAGVDFFVSTVDPLKEPPLITANTVLSILAVDYP 283
              || |:|||:::|  :|  :|  |  | ||| |:|||||.||||||:|||| :|||||
Db        233 KWYPINRETYVDRLIARY-GDGEESGLAPVDFFVSTVDPLKEPPLITANTVLSILAVDYP 291

Qy        284 VDKVSCYVSDDGAAMLSFESLVETAEFARKWVPFCKKFSIEPRAPEFYFSQKIDYLKDKV 343
              |:|:||||||||.:|| :|||| :|||:|:|||:|||:|||||||||||||:|||||||
Db        292 VEKISCYVSDDGSAMLTFESLAETAEYARKWVPFCKKYAIEPRAPEFYFSQKIDYLKDKI 351

Qy        344 QPSFVKERRAMKRDYEEYKVRVNALVAKAQKTPEEGWTMQDGTPWPGNNTRDHPGHDSGL 403
              |||||||||:|||||||||:|||||:|||:|| |:|||||||| ||||
Db        352 HPSFVKERRAMKRDYEEYKVRINALVAKAQKTPDEGWIMQDGTPWPGNNPRDHPGMIQVF 411

Qy        404 PWEILGARDIEGNELPRLVYVSREKRPGYQHHKKAGAENALVRVSAVLTNAPYILNVDCD 463
              |  |||| :|||||:|||||||||:|||||||| |||||||||||||:||||:|:|
Db        412 LGE-TGARDFDGNELPRLVYVSREKRPGYQHHKKAGAMNALVRVSAVLTNAPYILNLDCD 470

Qy        464 HYVNNSKAVREAMCFLMDPQVGRDVCYVQFPQRFDGIDKSDRYANRNVVFFDVNMKGLDG 523
              |||||||||||| :|| ||||||||||||:||||:|||||||||||||||||:||||:|:
Db        471 HYVNNSKAVREAMCFMMDPTVGRDVCYVQFPQRFDGIDRSDRYANRNVVFFDVNMKGLDG 530

Qy        524 IQGPVYVGTGCVFNRQALYGYGPPSMPSLRKRKDSSSCFSCCCPSKKKPAQDPAEVYRDA 583
              :|||||:|||| ||| :|||||||:|:| |   | |  ||| !|| :  |: ':
Db        531 LQGPVYVGTGCCFNRQALYGYGPPSLPALPKSSICSWC--CCCCPKKKVERSEREINRDS 588

Qy        584 KREDLNAAIFNLTEIDNYDEHERSMLISQLSFEKTFGLSSVFIESTLMENGGVPESANSP 643
              :|||| :|||| |||:|||:|:|||||| :|||:|||||'|||||||||||:|||||
Db        589 RREDLESAIFNLREIDNYDEYERSMLISQMSFEKSFGLSSVFIESTLMENGGVPESANPS 648

Qy        644 PFIKEAIQVIGCGYEEKTEWGKQIGWIYGSVTEDILSGFKMHCRGWRSIYCMPVRPAFKG 703
              |||| || |||| |||||:|:|||||||||||:|||||||||||||||||||||||||||
Db        649 TLIKEAIHVISCGYEEKTEWGKEIGWIYGSVTEDILTGFKMHCRGWRSIYCMPVRPAFKG 708

Qy        704 SAPINLSDRLHQVLRWALGSVEIFFSRHCPLWYGFGGGRLKWLQRLAYINTIVYPFTSLP 763
              |||||||||||||||| || |||:|||:|||:|||||||||:|||||||||||||||||||
Db        709 SAPINLSDRLHQVLRWALVSVEIFFSRHCPLWYGYGGGRLKWLQRLSYINTIVYPFTSLP 768

Qy        764 LIAYCTIPAVCLLTGKFIIPTLSNLASMLFLGLFISIIVTAVLELRWSGVSIEDLWRNEQ 823
              |:||| :||:||||||:|||||| |:: |||||:|||||:|||||||: ||| ||||
Db        769 LVAYCCLPAICLLTGKFIIPTLSNAATIWFLGLFMSIIVTSVLELRWSGIGIEDWWRNEQ 828

Qy        824 FWVIGGVSAHLFAVFQGFLKMLAGIDTNFTVTAKAAEDAEFGELYMVKWTTLLIPPTTLL 883
              ||||||||||||||:| |||:||:||.||||||:| :| :||||||: ||||:||||::|
```

```
Db        829 FWVIGGVSAHLFAVFQGILKMIAGLDTNFTVTAKATDDTEFGELYLFKWTTVLIPPTSIL 888

Qy        884 IINMSG-CAGFSDALNKGYEAWGPLFGKVFFAFWVILHLYPFLKGLMGRQNLTPTIVVLW 942
              ::|:    |||| ||| |: :||| | ||| | |||:|||||, |||||| ||||||.|
Db        889 VLNLVGVVAGFSAALNSGYESWGPLFGKVFFAMWVIMHLYPFLKGLMGRQNRTPTIVVLW 948

Qy        943 SVLLASVFSLVWVKINPFVNKVDNTLVAETC 973
              |||:|:|||:|| ::||   : |:
Db        949 SVLLASVFSLLWVKIDPFVGGTE TVNTNNT 979


RESULT 3
US-10-209-059-26
; Sequence 26, Application US/10209059
; Publication No. US20030163838A1
; GENERAL INFORMATION:
;  APPLICANT: Chugga, Kanwarpal S.
;  APPLICANT:  Wang, Haiyin
;  TITLE OF INVENTION: Maize Cellulose Synthases and Uses
;  TITLE OF INVENTION:  Thereof
;  FILE REFERENCE: 0864R2
;  CURRENT APPLICATION NUMBER: US/10/209,059
;  CURRENT FILING DATE:  2002-07-31
;  PRIOR APPLICATION NUMBER: 60/096,822
;  PRIOR FILING DATE: 1998-08-17
;  PRIOR APPLICATION NUMBER: 09/371,393
;  PRIOR FILING DATE: 1999-08-06
;  PRIOR APPLICATION NUMBER: 09/550,483
;  PRIOR FILING DATE: 2000-04-14
;  NUMBER OF SEQ ID NOS: 52
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 26
;   LENGTH: 1078
;   TYPE: PRT
;   ORGANISM: Zea mays
US-10-209-059-26

  Query Match           65.0%;  Score 3404;  DB 12;  Length 1078;
  Best Local Similarity   63.1%;  Pred. No. 2e-317;
  Matches  665;  Conservative  115;  Mismatches  174;  Indels  100;  Gaps   21;

Qy          9 CHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEINEGRKVCLRCGSPY ---- ---- 58
              | | ::|| :|: |||| || :|:|: |:|:| :|| : : :| : |
Db         41 CPVCADEVGTREDGQPFVACAECGFPVCRPCYEYERSEGTQCCPQCNTRYKRQKGCPRVE 100

Qy         59 -DEN  LLDDVEKKGSGNQSTMASHLNDQIQV  GIH  -  - -ARHISSVS 98
               :|| | :  :|  |  | :|  |  |:
Db        101 GDEEEGDEMDDFEDDFPA-KSRKNPMKLPVAIFVTMNPNKDFLGLWF GGQTLSSFTGSVA 159

Qy         99 TVDGEMNDEY-GNPIWKNRVESCKDKEKNRKPRSIKAETCALVPCEQQMEEKPSAEAGE 157
              | |  |:  ||:|.  | |: :| | :  |  :|  |
Db        160 GKDLEAEREMEGSMEWKDRIDKWKTKQEKRGKLNHDDSDE  KNEDEYM- -LLAEARQ 216

Qy        158 PLSIVYPIPRNKLTPYRAVIIMRLVILGLFFHFRITNPVESAFGLWLTSVICEIWFAFSW 217
              :  |||  : :|| ||::|||:  |  : | |  |  | | |||:||||||
Db        217 PLWRKVPIPSSMINPYRIVIVLRLVVLCFFLKFRITTPAIDAVPLWLASVICELWFAFSW 276

Qy        218 VLDQFPKWNPVNRETYIERLSARYEREGEPSQLAGVDFFVSTVDELKEPPLITANTVLSI 277
              :|| || || |||:::||. ||:|: |  :|: :||| ||| :||||:||||||||
Db        277 ILDQLPKWAPVTRETYLDRLALRYDREGEACRLSPIDFFVSTVDPLKEPPIITANTVLSI 336

Qy        278 LAVDYPVDKVSCYVSDDGAAMLSFESLVETAEFARKWVPFCKKFSIEPRAPEFYFSQKID 337
              |||:||:|||||||||:|  |::| |||||  :||||||||:|||||||||||||
Db        337 LAVDYPVDRVSCYVSDDGASMLLFDALSETAEFARRWVPFCKKFAVEPRAPEFYFSQKID 396

Qy        338 YLKDKVQPSFVKERRAMKREYEEFKVRVNALVAKAQKTPEEGWTMQDGTPWPGNNTRDHP 397
              |||. |||:|||||||:||:|::|||:||||.|||  |:|||:||||:|||||||||
Db        397 YLKDKVQPTFVKERRAMKREYEEFKVRINALVAKAQKKPEEGWVMQDGTPWPGNNTRDHP 456

Qy        398 GHDSGLPWEIL-------GAMDIEGNELPRLVYVSREKRPGYQHHKKAGAENALVRVSAV 450
              |       ::       |  :||:' |||  |||:|||||  |||||| ||||||''
Db        457 G--  ----MIQVYLGNQGALDVEGHELPRLVYVSREKRPGYDHHKKAGAMNALVRVSAV 508

Qy        451 LTNAPYILNVDCDHYVNNSKAVREAMCILMDPQVGRDVCYVQFPQRFDGIDKSDRYANRN 510
              |||||:|||:|||||||||||||||| |||||:|: :|||||||||||||: ||||||
Db        509 LTNAPFILNLDCDHYVNNSKAVREAMCFLMDPQLGKKLCYVQFPQRFDGIDRHDRYANRN 568

Qy        511 VVFFDVNMKGLDGIQGPVYVGTGCVFNRQALYGYGPPSMPSLRKRKD----SSSCFSCCC 566
              |||||:|||||||||||||||||||||||||| ||  |  | :       | |  |||
Db        569 VVFFDINMKGLDGIQGPVYVGTGCVFNRQALYGYDPP-RPEKRPKMTCDCWPSWCCCCCC 627

Qy        567 -------PSKKKPAQDPAE--------VYRDAKRED------------------LNAAIF 593
                    ::|    |  |          ||   ::|                    :   |
Db        628 FGGGKRGKARKDKKGDGGEEPRRGLLGFYRKRSKKDKLGGGSVAGSKKGGGLYKKHQRAF 687

Qy        594 NLTEI----DNYDEHERSMLISQLSFEKTFGLSSVFIESTLMENGGVPESANSPP--FIK 647
                 : ||    : ||| ||| |:|| |||| || | ||| |||:|:||:|: | : |   ||
Db        688 ELEEIEEGLEGYDELERSSLMSQKSFEKRFGQSPVFIASTLVEDGGLPQGAAADPAALIK 747

Qy        648 EAIQVIGCGYEEKTEWGKQIGWIYGSVTEDILSGFKMHCRGWRSIYCMPVRPAFKGSAPI 707
              ||| || ||||:|||||:|||||||||||||:|||||||||:|:|| | ||||||||||
Db        748 EAIHVISCGYEEKTEWGKEIGWIYGSVTEDILTGFKMHCRGWKSVYCTPTRPAFKGSAPI 807

Qy        708 NLSDRLHQVLRWALGSVEIFFSRHCPLWYGFGGGRLKWLQRLAYINTIVYPFTSLPLIAY 767
              |||||||||||||||||| |||||| | : |||||||:| || |||||||||:|:||
Db        808 NLSDRLHQVLRWALGSVEIFMSRHCPLRYAY-GGRLKWLERFAYTNTIVYPFTSIPLIAY 866

Qy        768 CTIPAVCLLTGKFIIPTLSNLASMLFLGLFISIIVTAVLELRWSGVSIEDLWRNEQFWVI 827
              |||||||||||||||||:||||: |: ||:||| |:|||||||||||| ||||||||
Db        867 CTIPAVCLLTGKFIIPTLNNLASIWFIALFLSIIATSVLELRWSGVSIEDWWRNEQFWVI 926

Qy        828 GGVSAHLFAVFQGFLKMLAGIDTNFTVTAKAAEDA--EFGELYMVKWTTLLIPPTTLLII 885
              |||||||||||||||:| |:||:|||||||| |    ||:||: |||||:|||||:|:|
Db        927 GGVSAHLFAVFQGFLKVLGGVDTSFTVTSKAAGDEADAFGDLYLFKWTTLLVPPTTLIII 986

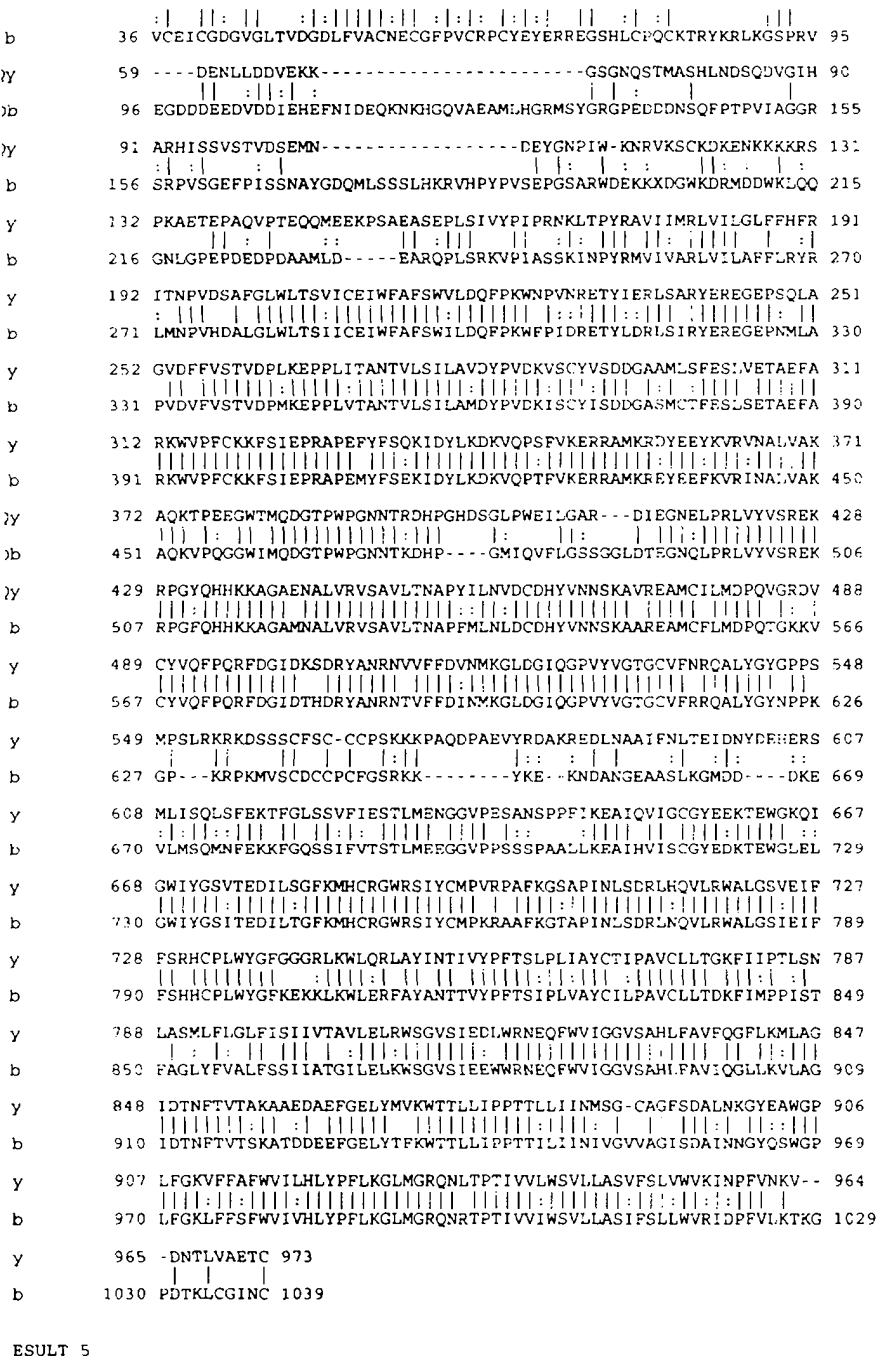
Qy        886 NMSG-CAGFSDALNKGYEAWGPLFGKVFFAFWVILHLYPFLKGLMGRQNLTPTIVVLWSV 944
              || |  || ||||:| || :|||||:||:|||:|||||||||||| |||||||:
Db        987 NMVGIVAGVSDAVNNGYGSWGPLFGKLFFSFWVIVHLYPFLKGLMGRQNRTPTIVVLWSI 1046

Qy        945 LLASVFSLVWVKINPFVNKVDNTLVAETCISIDC 978
              ||||:||||||:|:||: !    :: : |  ::|
Db       1047 LLASIFSLVWVRIDPFIPKAKGPIL-KPC-GVEC 1078


RESULT 4
US-09-900-237-14
; Sequence 14, Application US/09900237
; Patent No. US20020120124A1
; GENERAL INFORMATION:
;  APPLICANT: Allen, Stephen
;  TITLE OF INVENTION: Plant Cellulose Synthases
;  FILE REFERENCE: BB1170 US CIP
;  CURRENT APPLICATION NUMBER: US/09/900,237
;  CURRENT FILING DATE:  2001-07-06
;  PRIOR APPLICATION NUMBER: 60/092,844
;  PRIOR FILING DATE: 1998-07-14
;  PRIOR APPLICATION NUMBER: PCT/US99/15871
;  PRIOR FILING DATE: 1999-07-13
;  PRIOR APPLICATION NUMBER: 09/720383
;  PRIOR FILING DATE: 2000-12-21
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 14
;   LENGTH: 1039
;   TYPE: PRT
;   ORGANISM: Glycine max
;   FEATURE:
;   NAME/KEY: UNSURE
;   LOCATION: (201)
US-09-900-237-14

  Query Match           64.3%;  Score 3367;  DB 10;  Length 1039;
  Best Local Similarity   62.1%;  Pred. No. 1.5e-313;
  Matches  640;  Conservative  118;  Mismatches  182;  Indels   90;  Gaps   15;

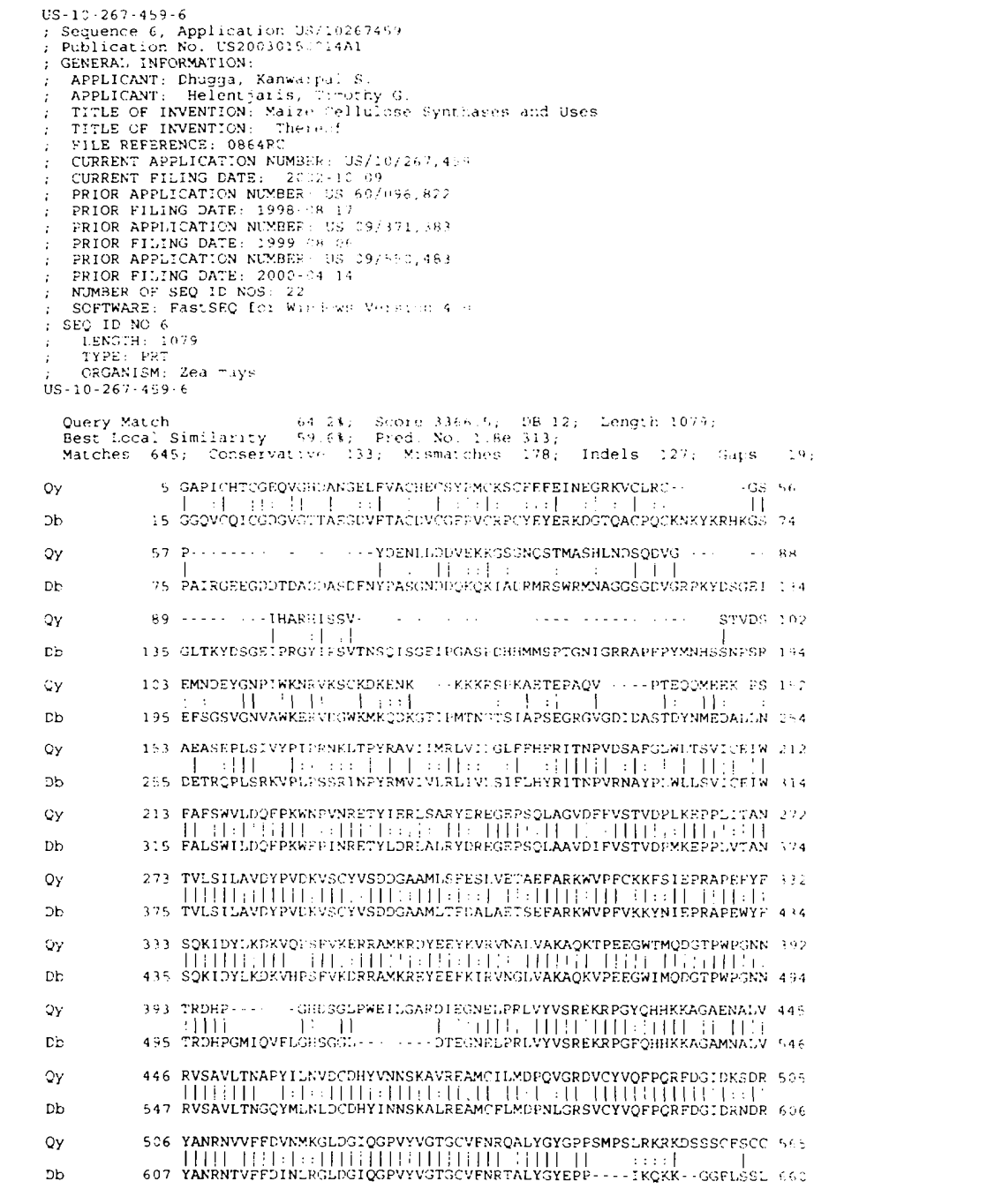
Qy          8 ICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEINEGRKVCLRC--------GSPY- 58
```

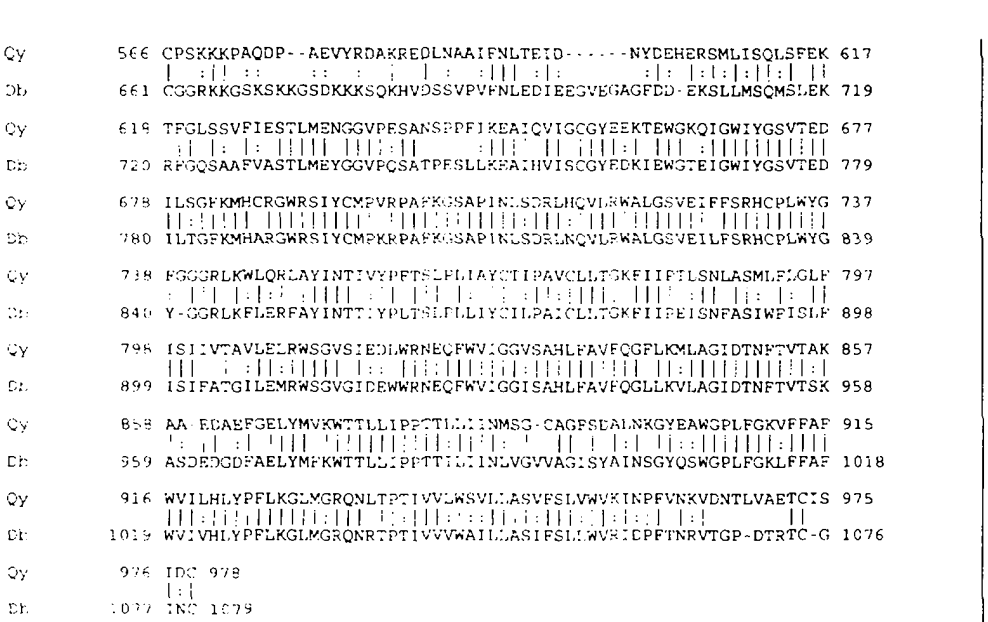


ESULT 5

US-10-267-459-6
; Sequence 6, Application US/10267459
; Publication No. US20030150014A1
; GENERAL INFORMATION:
; APPLICANT: Dhugga, Kanwarpal S.
; APPLICANT: Helentjaris, Timothy G.
; TITLE OF INVENTION: Maize Cellulose Synthases and Uses
; TITLE OF INVENTION: Thereof
; FILE REFERENCE: 0864PC
; CURRENT APPLICATION NUMBER: US/10/267,459
; CURRENT FILING DATE: 2002-10-09
; PRIOR APPLICATION NUMBER: US 60/096,822
; PRIOR FILING DATE: 1998-08-17
; PRIOR APPLICATION NUMBER: US 09/371,683
; PRIOR FILING DATE: 1999-08-09
; PRIOR APPLICATION NUMBER: US 09/550,483
; PRIOR FILING DATE: 2000-04-14
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 1079
; TYPE: PRT
; ORGANISM: Zea mays
US-10-267-459-6

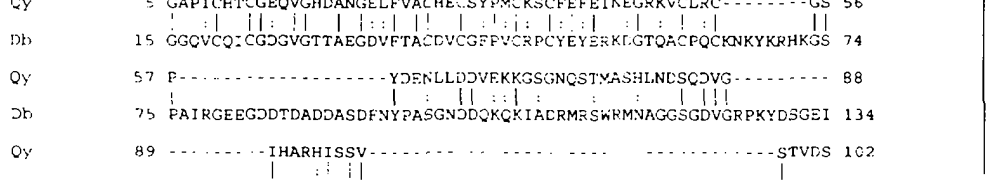
Query Match 64.2%; Score 3366.5; DB 12; Length 1079;
Best Local Similarity 59.6%; Pred. No. 1.8e-313;
Matches 645; Conservative 133; Mismatches 178; Indels 127; Gaps 19;

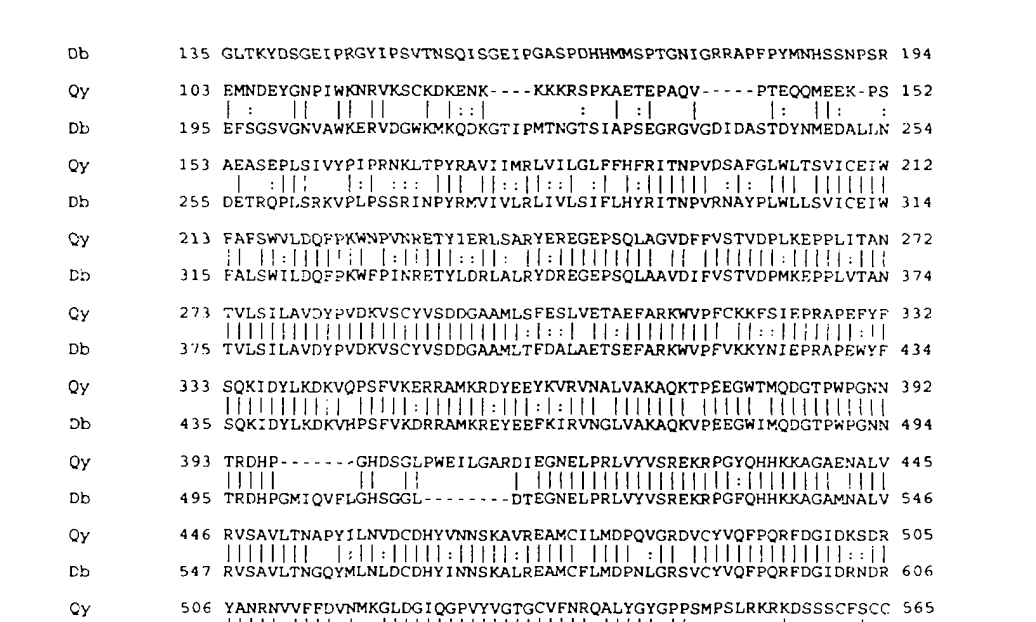




RESULT 6
US 10-209-059-22
; Sequence 22, Application US/10209059
; Publication No. US20030163838A1
; GENERAL INFORMATION:
; APPLICANT: Dhugga, Kanwarpal S.
; APPLICANT: Wang, Haiyin
; TITLE OF INVENTION: Maize Cellulose Synthases and Uses
; TITLE OF INVENTION: Thereof
; FILE REFERENCE: 0864R2
; CURRENT APPLICATION NUMBER: US/10/209,059
; CURRENT FILING DATE: 2002-07-31
; PRIOR APPLICATION NUMBER: 60/096,812
; PRIOR FILING DATE: 1998-08-17
; PRIOR APPLICATION NUMBER: 09/371,393
; PRIOR FILING DATE: 1999-08-06
; PRIOR APPLICATION NUMBER: 09/550,484
; PRIOR FILING DATE: 2000-04-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 22
; LENGTH: 1079
; TYPE: PRT
; ORGANISM: Zea mays
US 10-209-059-22

Query Match 64.2%; Score 3366.5; DB 12; Length 1079;
Best Local Similarity 59.6%; Pred. No. 1.8e-313;
Matches 645; Conservative 133; Mismatches 178; Indels 127; Gaps 19;





RESULT 7
US-10-209-059-10
; Sequence 10, Application US/10209059
; Publication No. US20030163838A1
; GENERAL INFORMATION:
; APPLICANT: Dhugga, Kanwarpal S.

APPLICANT: Wang, Haiyin
TITLE OF INVENTION: Maize Cellulose Synthases and Uses
TITLE OF INVENTION: Thereof
FILE REFERENCE: 0864R2
CURRENT APPLICATION NUMBER: US/10/209,059
CURRENT FILING DATE: 2002-07-31
PRIOR APPLICATION NUMBER: 60/096,822
PRIOR FILING DATE: 1998-08-17
PRIOR APPLICATION NUMBER: 09/371,383
PRIOR FILING DATE: 1999-08-06
PRIOR APPLICATION NUMBER: 09/550,483
PRIOR FILING DATE: 2000-04-14
NUMBER OF SEQ ID NOS: 52
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 10
LENGTH: 1077
TYPE: PRT
ORGANISM: Zea mays
-10-209-059-10

Query Match 64.2%; Score 3365.5; DB 12; Length 1077;
Best Local Similarity 59.7%; Pred. No. 2.3e-313;
Matches 645; Conservative 136; Mismatches 175; Indels 125; Gaps 20;

```
Qy          5 GAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEINEGRKVCLRC- ------GS 56
              |  :|  ||: ||  | |::| ||  | :|:|: |:|:|  :| : | :|        ||
Db         15 GGQVCQICGDGVGTTAEGDVFAACDVCGFPVCRPCYEYERKDGTQACPQCKTKYKRHKGS 74

Qy         57 P------------YDENLL----DDVEKKGSGNQSTMASHLNDSQDVG---------- 88
              |            | | |    :| ::| :   :   ::  | |||
Db         75 PAIRGEEGDDTDADSDFNYLASGNEDQKQKIADRMRSWRMNVGGSGDVGRPKYDSGEIGL 134

Qy         89 -------IHARHISSVS---------------------------------TVDSEM 104
                    |   :| ||:                                 |
Db        135 TKYDSGEIPRGYIPSVTNSQISGEIPGASPDHHMMSPTGNIGKRAPFPYVNHSPNPSREF 194

Qy        105 NDEYGNPIWKNRVKSCKDKENK----KKKRSPKAETEPAQV- ---PTEQQMEEK-PSAE 154
              :   || || ||   | |::|       :  | :|  |      |:  ||:   : |
Db        195 SGSIGNVAWKERVDGWKMKQDKGTIPMTNGTSIAPSEGRGVGDIDASTDYNMEDALLNDE 254

Qy        155 ASEPLSIVYPIPRNKLTPYRAVIIMRLVILGLFFHFRITNPVDSAFGLWLTSVICEIWFA 214
                :|||   |:| ::: ||| ||::||::| :| |:|||||| :|: ||| |||||||||
Db        255 TRQPLSRKVPLPSSRINPYRMVIVLRLIVLSIFLHYRITNPVRNAYPLWLLSVICEIWFA 314

Qy        215 FSWVLDQFPKWNPVNRETYIERLSARYEREGEPSQLAGVDFFVSTVDPLKEPPLITANTV 274
              ||:||||||| |:|||||::||: ||:|||||||| || |||||||:|||||:|||||
Db        315 LSWILDQFPKWFPINRETYLDRLALRYDREGEPSQLAAVDIFVSTVDPMKEPPLVTANTV 374

Qy        275 LSILAVDYPVDKVSCYVSDDGAAMLSFESLVETAEFARKWVPFCKKFSIEPRAPEFYFSQ 334
              |||||||||||||||||||||||||:|::| ||:|||||||| |||::|||||||:||||
Db        375 LSILAVDYPVDKVSCYVSDDGAAMLTFDALAETSEFARKWVPFVKKYNIEPRAPEWYFSQ 434

Qy        335 KIDYLKDKVQPSFVKERRAMKRDYEEYKVRVNALVAKAQKTPEEGWTMQDGTPWPGNNTR 394
              |||||||||  |||||:||||||:|||:||||| |||||:| |||||  |||||||||||||
Db        435 KIDYLKDKVHPSFVKDRRAMKREYEEFKVRVNGLVAKAQKVPEEGWIMQDGTPWPGNNTR 494

Qy        395 DHP-------GHDSGLPWEILGARDIEGNELPRLVYVSREKRPGYQHHKKAGAENALVRV 447
              |||       ||  ||          | ||||||||||||||||:|||||| ||||||
Db        495 DHPGMIQVFLGHSGGL--------DTEGNELPRLVYVSREKRPGFQHHKKAGAMNALVRV 546

Qy        448 SAVLTNAPYILNVDCDHYVNNSKAVREAMCILMDPQVGRDVCYVQFPQRFDGIDKSDRYA 507
              ||||||  |:||:|||||:|||||:|||| |||| :|: ||||||||||||||::||||
Db        547 SAVLTNGQYMLNLDCDHYINNSKALREAMCFLMDPNLGRSVCYVQFPQRFDGIDRNDRYA 606

Qy        508 NRNVVFFDVNMKGLDGIQGPVYVGTGCVFNRQALYGYGPPSMPSLRKRKDSSSCFSCCCP 567
              ||| ||||:|::|||||||||||||||||||| ||||| ||    ::::|      | |
Db        607 NRNTVFFDINLRGLDGIQGPVYVGTGCVFNRTALYGYEPP----IKQKK--GGFLSSLCG 660

Qy        568 SKKKPAQDP--AEVYRDAKREDLNAAIFNLTEID------NYDEHERSMLISQLSFEKTF 619
              :|| ::    ::  :  |  | :  :||| :|:       :|: |:|:|:||:| || |
Db        661 GRKKASKSKKGSDKKKSQKHVDSSVPVFNLEDIEEGVEGAGFDD-EKSLLMSQMSLEKRF 719

Qy        620 GLSSVFIESTLMENGGVPESANSPPFIKEAIQVIGCGYEEKTEWGKQIGWIYGSVTEDIL 679
```

```
Db        720 GQSAAFVASTLMEYGGVPQSATPESLLKEAIHVISCGYEDKTEWGTEIGWIYGSVTEDIL 779

Qy        680 SGFKMHCRGWRSIYCMPVRPAFKGSAPINLSDRLHQVLRWALGSVEIFFSRHCPLWYGFG 739
              :|||||  |||||:|||  |||:|:|||| :|||:|:|||:  |||||:|||  :|||||
Db        780 TGFKMHARGWRSIYCMPKRPAFKGSAPINLSDRLNQVLRWALGSVEILFSRHCPLWYGY- 838

Qy        740 GGRLKWLQRLAYINTIVYPFTSLPLIAYCTIPAVCLLTGKFIIPTLSNLASMLFLGLFIS 799
              |||:|:|: |||  :|| :|:|: || ::|||||||||| :|| ||: |: |||
Db        839 GGRLKFLERFAYINTIVYPLTSIPLIAYCILPAICLLTGKFIIPEISNFASIWFISLFIS 898

Qy        800 IIVTAVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQGFLKMLAGIDTNFTVTAKAA 859
                | :|:|:|||' | |:: ||.||||:  |:|||.|||||  ||:|||.|||||.::|||:
Db        899 IFATGILEMRWSGVSIDDWWRNEQFWVIGGISAHLFAVFQGLLKVLAGIDTNFTVTSKAS 958

Qy        860 -EDAEFGELYMVKWTTLLIPPTTLLIINMSG CAGFSDAINKGYEAWGPLFGKVFFAFWV 917
               || :| |||| |: ||| ||| :|:||: :  || | |:| :|::||||'|:||||'|'
Db        959 DRDGDFAELYMFKWTTLLIPPTTIIIINLVGVVAGISYAINSGYQSWGPLFGKLFFAFWV 1018

Qy        918 ILHLYPFLKGLMGRQNRTPTIVVLWSILLASVFSLVWVKINPFVNKVDNTLVAETCISID 977
              :||: |||  |||   :||  |:|  : |||:  |:|.|.:||  ::       :||  |:
Db       1019 IVHLYPFLKGLMGRQNRTPTIVVLWSILLASIFSLVWVRIDPFTTRVTGPDVEQCGINID 1078

Qy        978 C 978
              |
Db       1077 C 1077
```

RESULT 8
US-10-160-719-6
; Sequence 6, Application US/10160719
; Publication No. US20030167528A1
; GENERAL INFORMATION:
; APPLICANT: Kanwarpal S. Dhugga
; APPLICANT: Timothy G. Helentjaris
; APPLICANT: Benjamin A. Bowen
; APPLICANT: Xun Wang
; TITLE OF INVENTION: Maize Cellulose Synthases and Uses
; TITLE OF INVENTION: Thereof
; FILE REFERENCE: 0864
; CURRENT APPLICATION NUMBER: US/10/160,719
; CURRENT FILING DATE: 2002-06-03
; PRIOR APPLICATION NUMBER: US/09/371,383A
; PRIOR FILING DATE: 1999-08-06
; PRIOR APPLICATION NUMBER: US 60/096,822
; PRIOR FILING DATE: 1998-08-17
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 1077
; TYPE: PRT
; ORGANISM: Zea mays
US-10-160-719-6

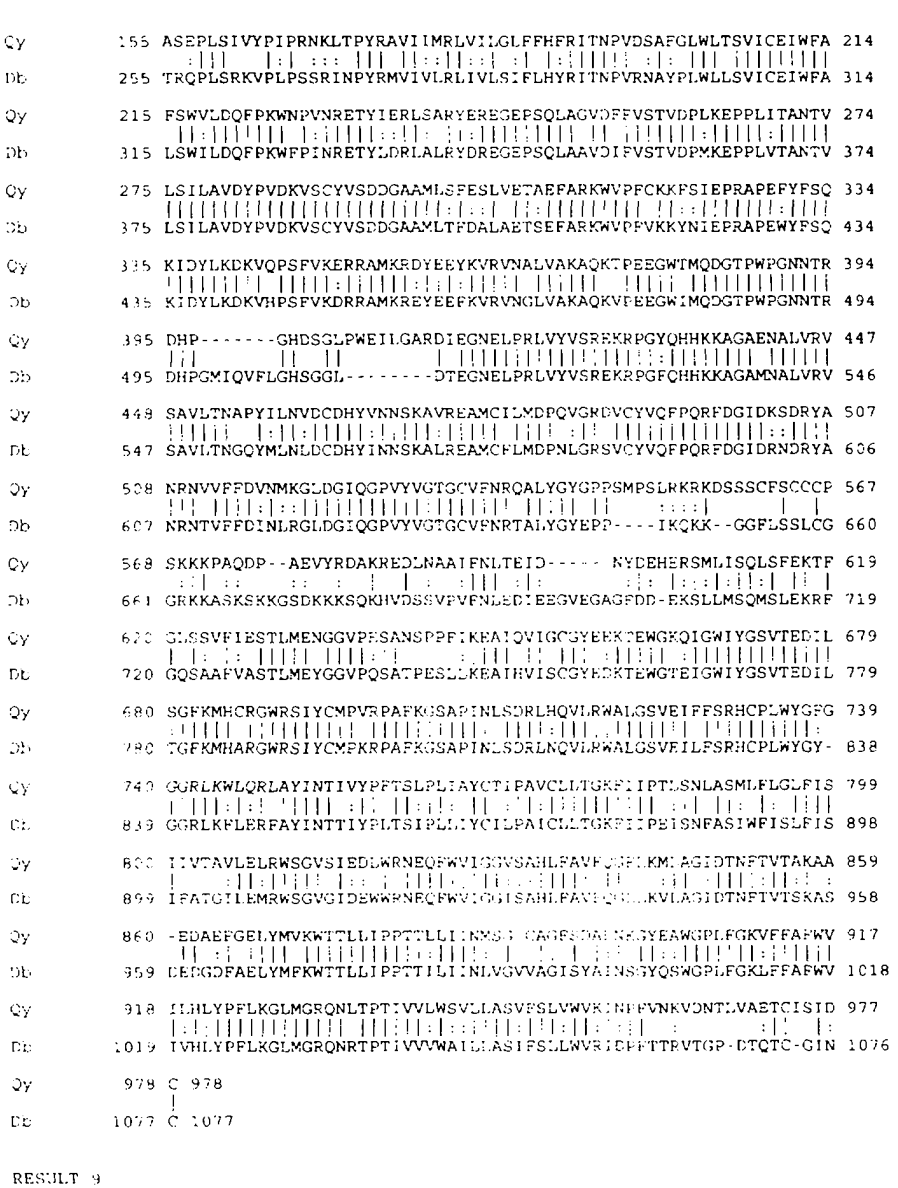
Query Match 64.2%; Score 3365.5; DB 12; Length 1077;
Best Local Similarity 59.7%; Pred. No. 2.3e-313;
Matches 645; Conservative 136; Mismatches 175; Indels 125; Gaps 20;

```
Qy          5 GAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEINEGRKVCLRC--------GS 56
              |  :|  ||: ||  | |::| ||  | :|:|: |:|:|  :| : | :|        ||
Db         15 GGQVCQICGDGVGTTAEGDVFAACDVCGFPVCRPCYEYERKDGTQACPQCKTKYKRHKGS 74

Qy         57 P------------YDENLL----DDVEKKGSGNQSTMASHLNDSQDVG---------- 88
              |            | | |    :| ::| :   :   ::  | |||
Db         75 PAIRGEEGDDTDADSDFNYLASGNEDQKQKIADRMRSWRMNVGGSGDVGRPKYDSGEIGL 134

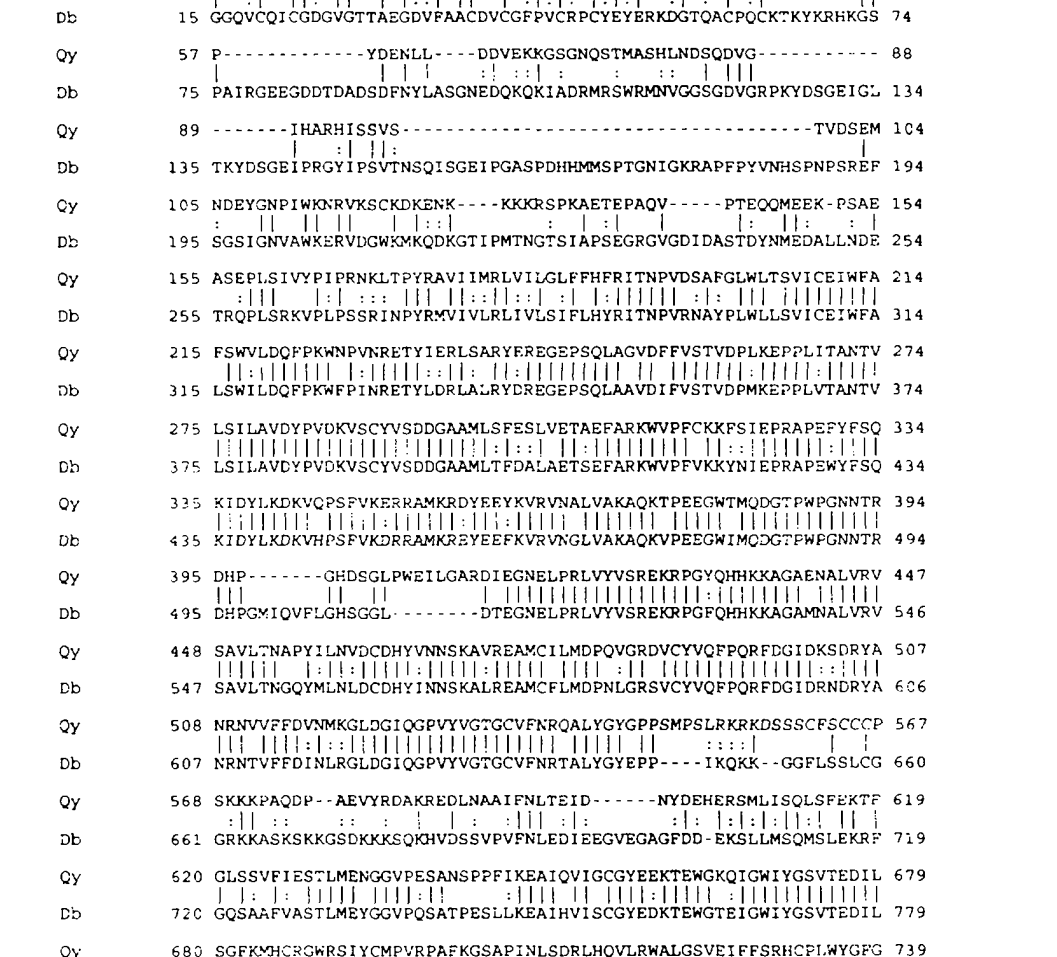
Qy         89 -------IHARHISSVS---------------------------------TVDSEM 104
                    |   :| ||:                                 |
Db        135 TKYDSGEIPRGYIPSVTNSQISGEIPGASPDHHMMSPTGNIGKRAPFPYVNHSPNPSREF 194

Qy        105 NDEYGNPIWKNRVKSCKDKENK----KKKRSPKAETEPAQV-----PTEQQMEEK-PSAE 154
              :   || || ||   | |::|       :  | :|  |      |:  ||:   : |
Db        195 SGSIGNVAWKERVDGWKMKQDKGTIPMTNGTSIAPSEGRGVGDIDASTDYNMEDALLNDE 254
```



RESULT 9
US-10-160-719-30
; Sequence 30, Application US/10160719
; Publication No. US20030167528A1
; GENERAL INFORMATION:
; APPLICANT: Kanwarpal S. Dhugga
; APPLICANT: Timothy G. Helentjaris
; APPLICANT: Benjamin A. Bowen
; APPLICANT: Xun Wang
; TITLE OF INVENTION: Maize Cellulose Synthases and Uses
; TITLE OF INVENTION: Thereof
; FILE REFERENCE: 0864
; CURRENT APPLICATION NUMBER: US/10/160,719
; CURRENT FILING DATE: 2002-06-03
; PRIOR APPLICATION NUMBER: US/09/371,383A
; PRIOR FILING DATE: 1999-08-06
; PRIOR APPLICATION NUMBER: US 60/096,822
; PRIOR FILING DATE: 1998-08-17
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 30
; LENGTH: 1077
; TYPE: PRT
; ORGANISM: Zea mays
US-10-160-719-30

Query Match 64.2%; Score 3365.5; DB 12; Length 1077;
Best Local Similarity 59.7%; Pred. No. 2.3e-313;
Matches 645; Conservative 136; Mismatches 175; Indels 125; Gaps 20;



```
       780 TGFKMHARGWRSIYCMPKRPAFKGSAPINLSDRLNQVLRWALGSVEILFSRHCPLWYGY- 838

       740 GGRLKWLQRLAYINTIVYPFTSLPLIAYCTIPAVCLLTGKFIIPTLSNLASMLFLGLFIS 799
           |||||:|:| ||||| :|| ||:||: || :||:||||||||||| :|| ||: |: ||||
       839 GGRLKFLERFAYINTTIYPLTSIPLLIYCILPAICLLTGKFIIPEISNFASIWFISLFIS 898

       800 IIVTAVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQGFLKMLAGIDTNFTVTAKAA 859
           |  | :||:||||| |:: |||||||||||:|||||||||| ||:|||||:||||:||:
       899 IFATGILEMRWSGVGIDEWWRNEQFWVIGGISAHLFAVFQGLLKVLAGIDTNFTVTSKAS 958

       860 -EDAEFGELYMVKWTTLLIPPTTLLIINMSG-CAGFSDALNKGYEAWGPLFGKVFFAFWV 917
            || :| |||| ||||||||||:||||: |  || | |:| ||::||:|||:||||: |
       959 DEDGDFAELYMFKWTTLLIPPTTILIINLVGVVAGISYAINSGYQSWGPLFGKLFFAFWV 1018

       918 ILHLYPFLKGLMGRQNLTPTIVVLWSVLLASVFSLVWVKINPFVNKVDNTLVAETCISID 977
           |:|||||||||||||| ||||||:|::|||||:|||:||:|:|| :|      :|| |:
      1019 IVHLYPFLKGLMGRQNRTPTIVVVWAILLASIFSLLWVRIDPFTTRVTGP-DTQTC-GIN 1076

       978 C 978
           |
      1077 C 1077


SULT 10
-10-209-059-14
Sequence 14, Application US/10209059
Publication No. US20030163838A1
GENERAL INFORMATION:
 APPLICANT: Dhugga, Kanwarpal S.
 APPLICANT:  Wang, Haiyin
 TITLE OF INVENTION: Maize Cellulose Synthases and Uses
 TITLE OF INVENTION:  Thereof
 FILE REFERENCE: 0864R2
 CURRENT APPLICATION NUMBER: US/10/209,059
 CURRENT FILING DATE:  2002-07-31
 PRIOR APPLICATION NUMBER: 60/096,822
 PRIOR FILING DATE: 1998-08-17
 PRIOR APPLICATION NUMBER: 09/371,383
 PRIOR FILING DATE: 1999-08-06
 PRIOR APPLICATION NUMBER: 09/550,483
 PRIOR FILING DATE: 2000-04-14
 NUMBER OF SEQ ID NOS: 52
 SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 14
  LENGTH: 1076
  TYPE: PRT
  ORGANISM: Zea mays
3-10-209-059-14

Query Match             64.2%;  Score 3362.5;  DB 12;  Length 1076;
Best Local Similarity   59.1%;  Pred. No. 4.4e-313;
Matches  638;  Conservative  145;  Mismatches  169;  Indels  127;  Gaps   19;

y           8 ICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEINEGRKVCLRC--------GSP-- 57
              :|  ||: ||  |:|:|| ||  | :|:|: |:|:|  :| : ; :|         |||
b          17 VCQICGDGVGTAADGDLFTACDVCGFPVCRPCYEYERKDGTQACPQCKTKYKRHKGSPPV 76

y          58 -------YDENLLDDVEKKGSGNQ------------------STMASHLNDSQDVGIHA 91
                     | : : |   : ||||                  | :   '' ::| |
b          77 HGEENEDVDADDVSDYNYQASGNQDQKQKIAERMLTWRTNSRGSDIGLAKYDSGEIG-HG 135

y          92 RH----------------------------ISSVSTV------ ---------DSEMN 105
              ::                          :| |  :                  | :
b         136 KYDSGEIPRGYIPSLTHSQISGEIPGASPDHMMSPVGNIGRRGHQFPYVNHSPNPSREFS 195

y         106 DEYGNPIWKNRVKSCKDKENKKKKRSPKAETEPAQ--------VPTEQQMEEK-PSAEAS 156
                || || ||   | |:      :      |::          |:  ||:   : |
b         196 GSLGNVAWKERVDGWKMKDKGAIPMTNGTSIAPSEGRGVADIDASTDYNMEDALLNDETR 255

y         157 EPLSIVYPIPRNKLTPYRAVIIMRLVILGLFFHFRITNPVDSAFGLWLTSVICEIWFAFS 216
              :|||   ||| ::: ||| ||::|| :| :|  :|||:||::|: ||| ||||||||| |
b         256 QPLSRKVPIPSSRINPYRMVIVLRLAVLCIFLRYRITHPVNNAYPLWLLSVICEIWFALS 315
```

```
Qy        217 WVLDQFPKWNPVNPETYIEREISARYFRDGPPSQLASVDFFVSTVDPLKEPPLITANTVLS 276
              |:|':||| :|:|,: |::' : |.. ||| ||| : |||,|':||||,:||||,:|
Db        316 WILDQFPKWSPINPETYLDRIALRYIRDGPPSQLAPVDIFVSTVDPMKEPPLVTANTVLS 375

Qy        277 ILAVDYPVDKVSCYVSDDGAAMLSFESLVETAEFARKWVPFCKKFSIEPRAPEFYFSQKI 336
              ;|||||:|||:|||:  ||| :|:  ||:::|||||||||:|||:||||:|:|||
Db        376 ILAVDYPVDKVSCYVSDDGAAMLTFEALSETSEFARKWVPFCKKYNIEPRAPEWYFAQKI 435

Qy        337 DYLKDKVQPSFVKERAMKRDYEEYKVRVNALVAKAQKTPEEGWTMQDGTPWPGNNTRDH 396
              ||:|||:  ||| |||:::||:: ||:  '||||:' |||| ||||:|||||' ||
Db        436 DYLKDKVQTNFVKERRAMKREYEEFKVRINGLVAKAQKVPEEGWIMQDGTPWPGNNTRDH 495

Qy        397 P---  --GPLSGLDPWETLGAFDIENELPRLVYVSREKRPGYQHHKKAGAENALVRVSA 449
                :       || :|         :||:|||| :|||||.|:||||||| ':| ||||'
Db        496 PGMIQVFLGHSGGL       --DVEGNELPRLVYVSREKRPGFQHHKKAGAMNALVRVSA 547

Qy        450 VLTNAPYILNVDCDHYVNNSKAVREAMCFMMDPQVGRDVCYVQFPQRFDGIDRKDRYANR 509
              |||| :|'|:|: ||:::||:' ||| :|| :'|||||||||:'||:::'|||
Db        548 VLTNGQYMLNLDCDHYINNSKAIREAMCFMMDPNLGRNVCYVQFPQRFDGIDRNDRYANR 607

Qy        510 NVVFFDVNMK-GLDGIQGPVYVGTGCVFNRTALYGYEPP--- VKKKK--PGFFSSLCGGR 563
              ||' :|:: ||| :|  ||  ||  ' | |  ::;::  :'  | :
Db        608 NTVFFDINLF-GLDGIQGPVYVGTGCVFNRTALYGYEPP--- VKKKK--PGFFSSLCGGR 661

Qy        564 KKPAQD--PAEVYSAAFEDLNAAIPNLTEID--  -NYDEHERSMLISQLSFEKTFGL 621
              ||::   :|  :  : |  :|||  :|:       :|: |:|:::|:|: ', ||
Db        662 KKTSKSKKSSRKKKSHRHADSGVPVFNLEDIEEGIEGSQFDD-EKSLIMSQMSLEKRFGQ 720

Qy        622 SSVFIESTLMENGGVPESANSPFFIKEAIQVIGCGYEEKTEWGKQIGWIYGSVTEDILSG 681
              ||||: ||  | |  :||     : ||| || ||||: |:|| :|||||:|||| :,
Db        721 SSVFVASTLMEYGNVPQSATPESLLKEAIHVISCGYEDKTDWGTEIGWIYGSVTEDILTG 780

Qy        682 FKMHCRGWRSIYCMPKRPAFKGSAPINLSDRLNQVLRWALGSVEIFFSRHCPLWYGFGGG 741
              |||| ||||,|':||  |||:::|||' ||||:|||||| |:|| |||||:'||. '|
Db        781 FKMHARGWRSIYCMPKRPAFKGSAPINLSDRLNQVLRWALGSIEILFSRHCPIWYGY-GG 839

Qy        742 RLKWLQRLAYINTIVYPFTSLPLIAYCTIPAVCLLTGKFIIPTLSNLASMLFLGLFISII 801
              |||:|:: ||||| :|| .|:|,: || .:|||| |.||||, :|||  : |: |.||
Db        840 RLKFLERFAYINTTIYPFTSIPLLIYCILPAVCLLTGKFIIPKISNLFSVWFISLFISIF 899

Qy        802 VTAVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQGFLKMLAGIDTNFTVTAKAAE 860
              | :||:|||'| |:. || ||||,'':|:|:'||||'| ||:|||||:|:'|:|;  '
Db        900 ATGILEMRWSGVGIDEWWRNEQFWVIGGISAHLFAVFQGLLKVLAGIDTSFTVTSKATE 959

Qy        861 DAEFGELYMVKWTTLLIPPTTLLIINMSG-CAGFSDALNKGYEAWGPLFGKVFFAFWVIL 919
              : :| ' || :|||:  |||:.|||:    ||   |:| | ::|||:|||:|:|.|||:
Db        960 EGDFAELYMFKWTTLLIPPTTILIINLVGVVAGISYAINSGYQSWGPLFGKLFFAFWVIV 1019

Qy        920 HLYPFLKGLMGRQNLTPTIVVLWSVLLASVFSLVWVKINPFVNKVDNTLVAETCISIDC 978
              ||:|||.:|||:: |||: |:|:: |||:|'|:|,,:||  :;    :|:    !:|
Db       1020 HLYPFLKGLMGKQNFTPTIVVVWAILLASIFSLMWVRIDPFTTRVTGPDIAK -CGINC 1076


RESULT 11
US-10-160-719 58
; Sequence 58, Application US/10160719
; Publication No. US20030167520A1
; GENERAL INFORMATION:
;  APPLICANT: Kanwarpal S. Dhugga
;  APPLICANT:  Timothy G. Helentjaris
;  APPLICANT:  Benjamin A   Bowen
;  APPLICANT:  Xun Wang
;  TITLE OF INVENTION: Maize Cellulose Synthases and Uses
;  TITLE OF INVENTION:  Thereof
;  FILE REFERENCE: 0864
;  CURRENT APPLICATION NUMBER: US/10/160,719
;  CURRENT FILING DATE:  2002-06-03
;  PRIOR APPLICATION NUMBER: US/09/371,383A
;  PRIOR FILING DATE: 1999-08-06
;  PRIOR APPLICATION NUMBER: US 60/096,822
;  PRIOR FILING DATE: 1998-08-17
;  NUMBER OF SEQ ID NOS: 60
;  SOFTWARE: FastSEQ for Windows Version 3.0
```

; SEQ ID NO 58
; LENGTH: 1076
; TYPE: PRT
; ORGANISM: Zea mays
US-10-160-719-58

Query Match 64.2%; Score 3362.5; DB 10; Length 1076;
Best Local Similarity 59.1%; Pred. No. 4.4e-313;
Matches 638; Conservative 145; Mismatches 169; Indels 127; Gaps 19;

```
Qy          8 ICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEINEGRKVCLRC--------GSP-- 57
              :|  ||: ||  |:|:|| |:  | .:|:|: ':|:|  :| : : .|        ||
Db         17 VCQICGDGVGTAADGDLFTACDVCGFPVCRPCYEYERKDGTQACPQCKTKYKRHKGSPPV 76

Qy         58     ----YDENLLDDVEKKGSGNQ-  ---     -----  - STMASHLNDSQDVGIHA 91
                    | : : |    : |||'                    | :      || ::| |
Db         77 HGEENEDVDADDVSDYNYQASGNQDQKQKIAERMLTWRTNGRGSDIGLAKYDSGEIG-HG 135

Qy         92 RH ----- ---- --------- -- - ---ISSVSTV-   -----------DSEMN 105
              ::                          :|  :                  | :
Db        136 KYDSGEIPRGYIPSLTHSQISGEIPGASPDHMMSPVGNIGFHGHQFPYVNHSPNPSREPS 195

Qy        106 DEYGNPIWKNRVKSCKDKENKKKKRSPKAETEPAQ----  --VPTEQQMEEK-PSAEAS 156
                ||  || ||   | |:      :      ::          |:  ||:    : |
Db        196 GSLGNVAWKERVDGWKMKDKGAIPMTNGTSIAPSEGRGVADIDASTDYNMEDALLNDETR 255

Qy        157 EPLSIVYPIPRNKLTPYRAVIIMRLVILGLFFHFRITNPVDSAFGLWLTSVICEIWFAFS 216
              :|||   ||| ::: ||| :|:::|| :| :|  :|||:||:::: |:  |||:|||| |
Db        256 QPLSRKVPIPSSRINPYRMVIVLRLAVLCIFLRYRITHPVNNAYPLWLLSVICEIWFALS 315

Qy        217 WVLDQFPKWNPVNRETYIERLSARYEREGEPSQLAGVDFFVSTVDPLKEPPLITANTVLS 276
              |:|,|||||:|:|||||::||: ::|||| ||| || ||||||:|||||:|||||||
Db        316 WILDQFPKWSPINRETYLDRLALRYDREGEPSQLAPVDIFVSTVDPMKEPPLVTANTVLS 375

Qy        277 ILAVDYPVDKVSCYVSDDGAAMLSFESLVETAEFARKWVPFCKKFSIEPRAPEFYFSQKI 336
              |||||||:|||||||||||||||:|::: ||:|| ||| :||||::|||||:|||||
Db        376 ILAVDYPVDKVSCYVSDDGAAMLTFDALSETSEFARKWVPFCKKYNIEPRAPEWYFAQKI 435

Qy        337 DYLKDKVQPSFVKERRAMKRDYEEYKVRVNALVAKAQKTPEEGWTMQDGTPWPGNNTRDH 396
              |:|'|||| |||:|||||||:|||:|:|  ||||| |||:  |||:|||||||||||:|
Db        436 DYLKDKVQTSFVKERRAMKREYEEFKVRINGLVAKAQKVPEEGWIMQDGTPWPGNNTRDH 495

Qy        397 P--------GHDSGLPWEILGARDIEGNELPRLVYVSREKRPGYQHHKKAGAENALVRVSA 449
              |        ||  ||        |:'|||||||||||||:|||||| ||||| ||
Db        496 PGMIQVFLGHSGGL---- ----DVEGNELPRLVYVSREKRPGFQHHKKAGAMNALVRVSA 547

Qy        450 VLTNAPYILNVDCDHYVNNSKAVREAMCFMMDPQVGRDV-VQFPQRFDGIDKSDRYANR 509
              |||| :|:|:|||||:|||| : |||| :|| : :|  |||||||||:::|:|||
Db        548 VLTNGQYMLNLDCDHYINNSKALREAMCFLMDPNIGRNV-VYVPFQGFDGIDRNDRYANR 607

Qy        510 NVVFFDVNMKGLDGIQGPVYVGTGCVFNRQALYGYGPPSMISLRKPKDSSSCFSCCCPSK 569
              | ||||:::|||||||||||||||:| ||:| ||      ::|:|      || |  :
Db        608 NTVFFDINLRGLDGIQGPVYVGTGCVFNRTALYGYEPP-   VKKKK--PGFFSSLCGGR 661

Qy        570 KKPAQD--PAEVYRDAKREDLNAAIFNLTEID---  --NYLEHERSMLISQLSFEKTFGL 621
              |: ::    :|  :  :  | :  :|||  :':       :| :|:|:::||:| || |:
Db        662 KKTSKSKKSSEKKKSHRHADSSVPVFNLEDIEEGIEGSQFDD-EKSLIMSQMSLEKRFGQ 720

Qy        622 SSVFIESTLMENGGVPESANSPPFIKEAIQVIGCGYEEKTEWGKQIGWIYGSVTEDILSG 681
              ||||: ||||| ||||:||      :|||| '  ||:.||:|' ':|||||||||||:|
Db        721 SSVFVASTLMEYGGVPQSATPESLLKEAIHVISCGYEDKTEWGTEIGWIYGSVTEDILTG 780

Qy        682 FKMHCRGWRSIYCMPVRPAFKGSAPINLSDRLHQVLRWALGSVEIFFSRHCPLWYGFGGG 741
              |||| |:|||||||:| ||||||||||||||:'::||:'|||:|:|| |||||:||: ||
Db        781 FKMHARGWRSIYCMPKRPAFKGSAPINLSDRLNQVLRWALGSIEILFSRHCPIWYGY-GG 839

Qy        742 RLKWLQRLAYINTIVYPFTSLPLIAYCTIPAVCLLTGKFIIPTLSNLASMLFLGLFISII 801
              ||:|:| |||:| :|| ||:||: || :|||||||||||:| :|:| |: |: |:|||
Db        840 RLKFLERFAYINTTIYPLTSIPLLLYCILPAVCLLTGKFIIPKISNLESVWFISLFISIF 899

Qy        802 VTAVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQGLLKVLAGIDTNFTVTAKAA-E 860
              :||:|||| |:: |||||||||:|||'|||||' ||:| |||:|||:||
Db        900 ATGILEMRWSGVGIDEWWRNEQFWVIGGISAHLFAVFQGLLKVLAGIDTSFTVTSKATDE 959

Qy        861 DAEFGELYMVKWTTLLIPPTTLLIINMSG-CAGFSDALNKGYEAWGPLFGKVFFAFWVIL 919
              : :| |||| ||||||||||:|||: |  || | |:| ||::|||||||:|||||:
Db        960 EGDFAELYMFKWTTLLIPPTTILIINLVGVVAGISYAINSGYQSWGPLFGKLFFAFWVIV 1019

Qy        920 HLYPFLKGLMGRQNLTPTIVVLWSVLLASVFSLVWVKINPFVNKVDNTLVAETCISIDC 978
              |||:|||||||:|| |||||:|:::|||:|||:||:|:||  :|    :|:    |:|
Db       1020 HLYPFLKGLMGKQNRTPTIVVVWAILLASIFSLMWVRIDPFTTRVTGPDIAK--CGINC 1076
```

RESULT 12
US-09-900-237-4
; Sequence 4, Application US/09900237
; Patent No. US20020120124A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Stephen
; TITLE OF INVENTION: Plant Cellulose Synthases
; FILE REFERENCE: BB1170 US CIP
; CURRENT APPLICATION NUMBER: US/09/900,237
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: 60/092,844
; PRIOR FILING DATE: 1998-07-14
; PRIOR APPLICATION NUMBER: PCT/US99/15871
; PRIOR FILING DATE: 1999-07-13
; PRIOR APPLICATION NUMBER: 09/720383
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 4
; LENGTH: 1148
; TYPE: PRT
; ORGANISM: Zea mays
US-09-900-237-4

Query Match 64.0%; Score 3352; DB 10; Length 1148;
Best Local Similarity 58.2%; Pred. No. 5.1e-312;
Matches 627; Conservative 153; Mismatches 178; Indels 120; Gaps 16;

```
Qy          1 MMESGAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEINEGRKVCLRCGSPY-- 58
              | :    :|  ||: || : :|| ||||:||::|:|: |:|:|  || : | :| : :
Db         90 MDQRNGQVCQICGDDVGRNPDGEPFVACNECAFPICRDCYEYERREGTQNCPQCKTRFKR 149

Qy         59 ----------DENLLDDVEKKGSGNQSTMASHLNDSQDVGIHARHISSVSTVD-----  101
                       :|: :||:| : :    : :| :|    :|| |:|     |
Db        150 LKGCARVPGDEEEDGVDDLENEFNWSDKHDSQYLAESM---LHA-HMSYGRGADLDGVPQ 205

Qy        102 ---- --------SEMNDE---------------------------------------  107
                         :| |:
Db        206 PFHPIPNVPLLTNGQMVDDIPPDQHALVPSFVGGGGKRIHPLPYADPNLPVQPRSMDPSK 265

Qy        108 ------YGNPIWKNRVKSCKDKENKKKKRSPKA---ETEPAQVPTEQQMEEKPSAEASEP 158
                   ||:  || |::| | |: :  :      : : | :|            || :|
Db        266 DLAAYGYGSVAWKERMESWKQKQERMHQTRNDGGGDDGDDADLPLMD--------EARQP 317

Qy        159 LSIVYPIPRNKLTPYRAVIIMRLVILGLFFHFRITNPVDSAFGLWLTSVICEIWFAFSWV 218
              ||   |:| ::: ||| :||:|||:|  |||:|: :||  || ||| |||||||| ||:
Db        318 LSRKIPLPSSQINPYRMIIIIRLVVLCFFFHYRVMHPVPDAFALWLISVICEIWFAMSWI 377

Qy        219 LDQFPKWNPVNRETYIERLSARYEREGEPSQLAGVDFFVSTVDPLKEPPLITANTVLSIL 278
              ||||||| |: ||||::||| |:::|| ||||| |||||||||||||||||:|||||||
Db        378 LDQFPKWFPIERETYLDRLSLRFDKEGHPSQLAPVDFFVSTVDPLKEPPLVTANTVLSIL 437

Qy        279 AVDYPVDKVSCYVSDDGAAMLSFESLVETAEFARKWVPFCKKFSIEPRAPEFYFSQKIDY 338
              :|||||||||||||||||||:||:| ||:|||:|||||||::|:||||:|| |||||
Db        438 SVDYPVDKVSCYVSDDGAAMLTFEALSETSEFAKKWVPFCKRYSLEPRAPEWYFQQKIDY 497

Qy        339 LKDKVQPSFVKERRAMKRDYEEYKVRVNALVAKAQKTPEEGWTMQDGTPWPGNNTRDHPG 398
              ||||| |:||:|||||||:|||:|||:|||||||| |||||||||||||||| |||||
Db        498 LKDKVAPNFVRERRAMKREYEEFKVRINALVAKAQKVPEEGWTMQDGTPWPGNNVRDHPG 557

Qy        399 HDSGLPWEILGARDIEGNELPRLVYVSREKRPGYQHHKKAGAENALVRVSAVLTNAPYIL 458
                    :  |  |:|||||||||||||||||| ||||||| |||||||||||||||:|
Db        558 MIQVFLGQ-SGGHDVEGNELPRLVYVSREKRPGYNHHKKAGAMNALVRVSAVLTNAPYLL 616
```

```
        459 NVDCDHYVNNSKAVREAMCILMDPQVGRDVCYVQFPQRFDGIDKSDRYANRNVVFFDVNM 518
        617 NLDCDHYINNSKAIKEAMCFMMDPLLGKKVCYVQFPQRFDGIDRHDRYANRNVVFFDINM 676

        519 KGLDGIQGPVYVGTGCVFNRQALYGYGPPSMPSLRKRKDSSS-------CFSCCC----- 566
        677 KGLDGIQGPIYVGTGCVFRRQALYGYDAPK----TKKPPSRTCNCWPKWCICCCCFGNRK 732

        567 PSKKKPAQDP--AEVYRDAKREDLNAAIFNLTEIDNY---DEHERSMLISQLSFEKTFGL 621
        733 TKKKTKTSKPKFEKIKKLFKKKENQAPAYALGEIDEAAPGAENEKASIVNQQKLEKKFGQ 792

        622 SSVFIESTLMENGGVPESANSPPFIKEAIQVIGCGYEEKTEWGKQIGWIYGSVTEDILSG 681
        793 SSVFVASTLLENGGTLKSASPASLLKEAIHVISCGYEDKTGWGKDIGWIYGSVTEDILTG 852

        682 FKMHCRGWRSIYCMPVRPAFKGSAPINLSDRLHQVLRWALGSVEIFFSRHCPLWYGFGGG 741
        853 FKMHCHGWRSIYCIPKRAAFKGSAPLNLSDRLHQVLRWALGSIEIFFSNHCPLWYGYGGG 912

        742 RLKWLQRLAYINTIVYPFTSLPLIAYCTIPAVCLLTGKFIIPTLSNLASMLFLGLFISII 801
        913 -LKFLERFSYINSIVYPWTSIPLLAYCTLPAICLLTGKFITPELNNVASLWFMSLFICIF 971

        802 VTAVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQGFLKMLAGIDTNFTVTAKAAED 861
        972 ATSILEMRWSGVGIDDWWRNEQFWVIGGVSSHLFAVFQGLLKVIAGVDTSFTVTSKGGDD 1031

        862 AEFGELYMVKWTTLLIPPTTLLIINMSG-CAGFSDALNKGYEAWGPLFGKVFFAFWVILH 920
       1032 EEFSELYTFKWTTLLIPPTTLLLLNFIGVVAGISNAINNGYESWGPLFGKLFFAFWVIVH 1091

        921 LYPFLKGLMGRQNLTPTIVVLWSVLLASVFSLVWVKINPFVNKVDNTLVAETCISIDC 978
       1092 LYPFLKGLVGRQNRTPTIVIVWSILLASIFSLLWVRIDPFLAKDDGPLLEE--CGLDC 1147


ESULT 13
S-09-900 237-26
 Sequence 26, Application US/09900237
 Patent No. US20020120124A1
 GENERAL INFORMATION:
  APPLICANT: Allen, Stephen
  TITLE OF INVENTION: Plant Cellulose Synthases
  FILE REFERENCE: BB1170 US CIP
  CURRENT APPLICATION NUMBER: US/09/900,237
  CURRENT FILING DATE:  2001-07-06
  PRIOR APPLICATION NUMBER: 60/092,844
  PRIOR FILING DATE: 1998-07-14
  PRIOR APPLICATION NUMBER: PCT/US99/15871
  PRIOR FILING DATE: 1999-07-13
  PRIOR APPLICATION NUMBER: 09/720383
  PRIOR FILING DATE: 2000-12-21
  NUMBER OF SEQ ID NOS: 33
  SOFTWARE: Microsoft Office 97
 SEQ ID NO 26
    LENGTH: 1091
    TYPE: PRT
    ORGANISM: Impatiens balsamia
S-09-900-237-26

 Query Match             63.9%;  Score 3348.5;  DB 10;  Length 1091;
 Best Local Similarity   59.8%;  Pred. No. 1e-311;
 Matches  643;  Conservative  135;  Mismatches  183;  Indels  115;  Gaps   22;

y          1 MMESGAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEINEGRKVCLRC------ 54
b         33 LKEANGQICQICGDTVGKSATGDTFVACNECGFPVCRPCYEYERKDGNQCCPQCKTRYKR 92

y         55 --GSP-----YDENLLDDVEKK----GSG-NQSTMAS----HLNDSQDVGI--------- 89
b         93 QKGSPRVEGDEEEEDVDDLENEFNYSGKGKNQKKVTTARRPWQGDQQDIELSVSSSRHDE 152

Qy         90 -------HARHIS-------------------SVSTVD------------SE 103
Db        153 SQQPVPLLTHGHSVR-EIPTPDNHSIPTTSGPISFVEKSIPYIDPRQPVAVRIIVDPSKD 212

Qy        104 MND-EYGNPIWKNRVESCKDKENFE----KKRSPKAETEPAQVPTEQQMEEKPSAEASEP 158
Db        213 LNSYGLGNVDWKERVEQWKLKQEKNMVQMTSRYPEGKGDTEGTGSNGEELQMAADDIRQP 272

Qy        159 LSIVYPIPRNKLTPYRAVIIMRLVILGLFFHFRITNPVDSAFGLWLTSVICEIWFAFSWV 218
Db        273 MSRIVPISSTHLTPYRVVIILRLIILGFFLQYRCTHPVKDAYPLWLTSVICEVWFALSWL 332

Qy        219 LDQFPKWNPVNRETYIERLSARYERDGEPSQLAGVDFFVSTVDPLKEPPLITANTVLSIL 278
Db        333 LDQFPKWSPVNRETYIERLSMRFDREGEPSQLAPIDVFVSTVDPLKEPPLVTANTVLSIL 392

Qy        279 AVDYPVDKVSCYVS[illegible]AAMLSFESLVETAAFAKWVPFCKKFSIEPRAPEFYFSQKIDY 338
Db        393 AVDYPVDKVSCYVS[illegible]AMLTFEALSETAEFAKRWAPFCKKHSIEPRAPEFYFAQKIDY 452

Qy        339 LKDKVQPSFVKERRAMKREYEEFKVRINALVAKAQKTPEEGWTMQDGTPWPGNNTRDHP  397
Db        453 LKDKVQFSFVKERRAMKREYEEFKVRINALVAKAQKVPEEGWTMQDGTPWPGNNSRDHPG 512

Qy        398 ------GHDSGLPWPLVGARDIEGNELPRLVYVSREKRPGYQHHKKAGAENALVRVSAVL 451
Db        513 MIQVFLGHSGGF---     -DTEGNELPRLVYVSREKRPGFQHHKKAGAMNALIRVSAVL 564

Qy        452 TNAPYILNVDCDHYVNNSKAVREAMCILMDPQVGRDVCYVQFPQRFDGIDKSDRYANPNV 511
Db        565 TNGAYLLNVDCDHYFNNSKCLKEAMCFMMDPNLGKKTCYVQFPQRFDGIDLHDRYANRNI 624

Qy        512 VFFDVNMKGLDGIQGPVYVGTGCVFNRQALYGYGPPSMPSLRKRKDSSSCFSCCCPSKKK 571
Db        625 VFFDINLKGLDGIQGPVYVGTGCCFNRQALYGYDP---VLTEEDLEPNIIIKSCCGSRKK 681

Qy        572 PA---QDPAEVYRDAKPEDLNAAIFNLTE    IDNYDEHERSMLISQLSFEKTFGLSSV 624
Db        682 GKGGNKKYIDKNRALKRTESTAPIFNMEDIEEGIEGYDD-ERSFLMAQ-SYEKRFGQSPV 739

Qy        625 FIESTLMENGGVPESANSPPFIKEAIQVIGCGYEEKTEWGKQIGWIYGSVTEDILSGFKM 684
Db        740 LIAATFMEQGGLPFSTNSATLLKEAIHVISCGYEDKTEWGKEIGWIYGSVTEDILTGFKM 799

Qy        685 HCRGWRSIYCMPVRPAFKGSAPINLSDRLHQVLRWALGSVEIFFSRHCPLWYGFGGGRLK 744
Db        800 HTRGWISIYCMPPRPAFKGSAPINLSDRLNQVLRWALGSIEILLSRHCPIWYGY-SGRLK 858

Qy        745 WLQRLAYINTIVYPFTSLPLIAYCTIPAVCLLTGKFIIPTLSNLASMLFLGLFISIIVTA 804
Db        859 FLERLAYINTIVYPLTSIPLLAYCTLPAICLLTGKFIVPEISNYASIWFILLFVSIFSTG 918

Qy        805 VLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQGFLKMLAGIDTNFTVTAKAA-EDAE 863
Db        919 ILELRWSGVTLEDWWRNEQFWVIGGTSAHLFAVFQGLLKVLAGIDTNFTVTSKASDEDGD 978

Qy        864 FGELYMVKWTTLLIPPTTLLIINMSG-CAGFSDALNKGYEAWGPLFGKVFFAFWVILHLY 922
Db        979 FAELYVFKWTSLLIPPTTILVVNMVGIVAGVSFAINSGYQSWGPLFGRLFFAIWVIVHLY 1038

Qy        923 PFLKGLMGRQNLTPTIVVLWSVLLASVFSLVWVKINPFVNKVDNTLVAETCISIDC 978
Db       1039 PFLKGLLGRQNRTPTIVIVWSVLLASIFSLLWVRIDPFTS--DSTKARGQC-GIDC 1091


RESULT 14
US-10-229-193-6
; Sequence 6, Application US/10229193
; Publication No. US20030126643A1
; GENERAL INFORMATION:
;  APPLICANT: Arioli, Antonio Williamson, Richard E.
;  APPLICANT:  Betzner, Andreas S.
```

```
;  APPLICANT:  Peng, Liangcai
;  TITLE OF INVENTION: Manipulation of Cellulose and/or Beta-1.4-Glucan
;  FILE REFERENCE: 96-98A
;  CURRENT APPLICATION NUMBER: US/10/229,193
;  CURRENT FILING DATE:  2002-08-26
;  PRIOR APPLICATION NUMBER: 09/221,013
;  PRIOR FILING DATE: 1998-12-23
;  PRIOR APPLICATION NUMBER: PCT/AU97/00402
;  PRIOR FILING DATE: 1997-06-24
;  PRIOR APPLICATION NUMBER: AU PO0699
;  PRIOR FILING DATE: 1996-06-27
;  NUMBER OF SEQ ID NOS: 36
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
;   LENGTH: 1081
;   TYPE: PRT
;   ORGANISM: Arabidopsis thaliana
US-10-229-193-6

  Query Match           63.9%;  Score 3347;  DB 15;  Length 1081;
  Best Local Similarity   61.3%;  Pred. No. 1.4e-311;
  Matches  644;  Conservative  122;  Mismatches  176;  Indels  108;  Gaps   18;

Qy          8 ICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEINEGRKVCLRC--------GSP-  57
              ||  ||: ||    |::||||:|:::|:|: |:|:|  :| : | :|         |||
Db         38 ICQICGDDVGLAETGDVFVACNECAFPVCRPCYEYERKDGTQCCPQCKTRFRRHRGSPRV 97

Qy         58 ---YDENLLDDVEKKGSGNQ-STMASHLNDSQDVGIHARHIS--------SVST 99
                 ||: :||:| : :  | :  | |    ::    :|' |           : |
Db         98 EGDEDEDDVDDIENEFNYAQGANKARHQRHGEEFSSSSRHESQPIPLLTHGHTVSGEIRT 157

Qy        100 VDSE   - - ---------MNDEY-------------------- GNPIWKNRVKSC 120
                |::              ::  |                    ||  || ||:
Db        158 PDTQSVRTTSGPLGPSDRNAISSPYIDPRQPVPVRIVDPSKDLNSYGLGNVDWKERVEGW 217

Qy        121 KDKENK-------KKKRSPKAETEPAQVPTEQQMEEKPSAEASE-PLSIVYPIPRNKLTP 172
              | |: |       |      | |     |   :|   |: :  |:| | ||| ::|||
Db        218 KLKQEKNMLQMTGKYHEGKGGEIE----GTGSNGEELQMADDTRLPMSRVVPIPSSRLTP 273

Qy        173 YRAVIIMRLVILGLFFHFRITNPVDSAFGLWLTSVICEIWFAFSWVLDQFPKWNPVNRET 232
              || |||:|:||  |  :| |:|| :|: ||||||||||||||||:||||||| |:||||
Db        274 YRVVIILRLIILCFFLQYRTTHPVKNAYPLWLTSVICEIWFAFSWLLDQFPKWYPINRET 333

Qy        233 YIERLSARYEREGEPSQLAGVDFFVSTVDPLKEPPLITANTVLSILAVDYPVDKVSCYVS 292
              |::||: |||:|||||,  || ;|||| |||:|':||  |||| :|||:||||:|||
Db        334 YLDRLAIRYDRDGEPSQLVPVDVFVSTVDPLKEPPLVTANTVLSILSVDYPVDKVACYVS 393

Qy        293 DDGAAMLSFESLVETAEFAKKWVPFCKKFNIEPRAPEFYFAQKIDYLKDKVQPSFVKERR 352
              |||:|||:|||| ||||||.||';||| :|||: |||  .|||:|||:|||||||||
Db        394 DDGSAMLTFESLSETAEFAKKWVPFCKKFNIEPRAPEFYFAQKIDYLKDKIQPSFVKERR 453

Qy        353 AMKRDYEEYKVRVNALVAKAQKTPEEGWTMQDGTPWPGNNTRDHPGMIQVFLGHSGGL-- 410
              |||:||:|||:|||||||| ||||| ||||'|||  |||| :        ||  ||
Db        454 AMKREYEEFKVRINALVAKAQKIPEEGWTMQDGTPWPGNNTRDHPGMIQVFLGHSGGL-- 511

Qy        411 EILGARDIEGNELPRLVYVSREKRPGYQHHKKAGAENALVRVSAVLTNAPYILNVDCDHY 470
              | :|||||:|||||:|:|'''|| ||| ||:|:|||:|||||:|||:|:|||||||||
Db        512 -  - -  DTDGNELPRLIYVSREKRPGFQHHKKAGAMNALIRVSAVLTNGAYLLNVDCDHY 565

Qy        471 VNNSKAVREAMCILMDPQVGRDVCYVQFPQRFDGIDKSDRYANRNVVFFDVNMKGLDGIQ 530
              ||||::|||| :||| ::  |,,'|||'|||  |||  ||:||||:|:||||||||
Db        566 FNNSKAIKEAMCFMMDPAIGKKCCYVQFPQRFDGIDLHDRYANRNIVFFDINMKGLDGIQ 625

Qy        531 GPVYVGTGCVFNRQALYGYGPPSMPSLRKRKDSSSCFSCCCPSKKKPAQDPAEVY---RD 587
              |:|||||| ||||'|||| |       :  : :      || ':||         |   |
Db        626 GPVYVGTGCCFNRQALYGYDP---VLTEEDLEFNIIVKSCCGSRKKGKSSKKYNYEKRRG 682

Qy        588 AKREDLNAAIFNLTEID----NYDEHERSMLISQLSFEKTFGLSSVFIESTLMENGGVPE 643
              | | || :||: :||     ||:  ||:,:|| |  | |  | |:: :| || ||:'
Db        683 INRSDSNAPLFNMEDIDEGFEGYDD-ERSILMSQRSVEKRFGQSPVFIAATFMEQGGIPP 741

Qy        644 SANSPPFIKEAIQVIGCGYEEKTEWGKQIGWIYGSVTEDILSGFKMHCRGWRSIYCMPVR 703
              : |    :|||| :' |||:: |||||:||| :||||:||:|:||| '|| :|||  : |
Db        742 TTNPATLLKEAIHVISCGYEDKTEWGKEIGWIYGSVTEDILTGFKMHARGWISIYCNPPR 801

Qy        704 PAFKGSAPINLSDRLHQVLRWALGSVEIFFSRHCPLWYGFGGGRLKWLQRLAYINTIVYP 763
              |||||||||||||||:|||||||||:||  |||||:|||:  |||: |:|:|||||||||
Db        802 PAFKGSAPINLSDRLNQVLRWALGSIEILLSRHCPIWYGY-HGRLRLLERIAYINTIVYP 860

Qy        764 FTSLPLIAYCTIPAVCLLTGKFIIPTLSNLASMLFLGLFISIIVTAVLELRWSGVSIEDL 823
              ||:|||||| :|| ||:|  :|||| :|| ||: |: ||||| || :||||||||||||
Db        861 ITSIPLIAYCILPAFCLITDRFIIPEISNYASIWFILLFISIAVTGILELRWSGVSIEDW 920

Qy        824 WRNEQFWVIGGVSAHLFAVFQGFLKMLAGIDTNFTVTAKAA-EDAEFGELYMVKWTTLLI 882
              |||||||||| ||||||||| ||:|||||||||||:||  || :| |||: ||| |||
Db        921 WRNEQFWVIGGTSAHLFAVFQGLLKVLAGIDTNFTVTSKATDEDGDFAELYIFKWTALLI 980

Qy        883 PPTTLLIINMSG-CAGFSDALNKGYEAWGPLFGKVFFAFWVILHLYPFLKGLMGRQNLTP 941
              ||||:|::|: |  || | |:| ||::|||||||:||| ||| |||||||||:||| ||
Db        981 PPTTVLLVNLIGIVAGVSYAVNSGYQSWGPLFGKLFFALWVIAHLYPFLKGLLGRQNRTP 1040

Qy        942 TIVVLWSVLLASVFSLVWVKINPFVNKVDN 971
              |||::|||||||:|||:||:|||||:   |
Db       1041 TIVIVWSVLLASIFSLLWVRINPFVDANPN 1070


RESULT 15
US-10-229-193-12
; Sequence 12, Application US/10229193
; Publication No. US20030126643A1
; GENERAL INFORMATION:
;  APPLICANT: Arioli, Antonio Williamson, Richard E.
;  APPLICANT:  Betzner, Andreas S.
;  APPLICANT:  Peng, Liangcai
;  TITLE OF INVENTION: Manipulation of Cellulose and/or Beta-1.4-Glucan
;  FILE REFERENCE: 96-98A
;  CURRENT APPLICATION NUMBER: US/10/229,193
;  CURRENT FILING DATE:  2002-08-26
;  PRIOR APPLICATION NUMBER: 09/221,013
;  PRIOR FILING DATE: 1998-12-23
;  PRIOR APPLICATION NUMBER: PCT/AU97/00402
;  PRIOR FILING DATE: 1997-06-24
;  PRIOR APPLICATION NUMBER: AU PO0699
;  PRIOR FILING DATE: 1996-06-27
;  NUMBER OF SEQ ID NOS: 36
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
;   LENGTH: 1081
;   TYPE: PRT
;   ORGANISM: Arabidopsis thaliana
US-10-229-193-12

  Query Match           63.8%;  Score 3343;  DB 15;  Length 1081;
  Best Local Similarity   61.2%;  Pred. No. 3.3e-311;
  Matches  643;  Conservative  122;  Mismatches  177;  Indels  108;  Gaps   18;

Qy          8 ICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEINEGRKVCLRC--------GSP-- 57
              ||  ||: ||    |::||||:|::|:|: |:|:|  :| : | :|         |||
Db         38 ICQICGDDVGLAETGDVFVACNECAFPVCRPCYEYERKDGTQCCPQCKTRFRRHRGSPRV 97

Qy         58 ---YDENLLDDVEKKGSGNQ-STMASHLNDSQDVGIHARHIS-------------SVST 99
                 ||: :||:| : :  | :  | |    ::    :|| |               : |
Db         98 EGDEDEDDVDDIENEFNYAQGANKARHQRHGEEFSSSSRHESQPIPLLTHGHTVSGEIRT 157

Qy        100 VDSE----------------MNDEY----------------------GNPIWKNRVKSC 120
                |::              ::  |                      ||  || ||:
Db        158 PDTQSVRTTSGPLGPSDRNAISSPYIDPRQPVPVRIVDPSKDLNSYGLGNVDWKERVEGW 217

Qy        121 KDKENK-------KKKRSPKAETEPAQVPTEQQMEEKPSAEASE-PLSIVYPIPRNKLTP 172
              | |: |       |      | |     |   :|   |: :  |:| | ||| ::|||
Db        218 KLKQEKNMLQMTGKYHEGKGGEIE----GTGSNGEELQMADDTRLPMSRVVPIPSSRLTP 273

Qy        173 YRAVIIMRLVILGLFFHFRITNPVDSAFGLWLTSVICEIWFAFSWVLDQFPKWNPVNRET 232
              || |||:|:||  |  :| |:|| :|: ||||||||||||||||:||||||| |:||||
Db        274 YRVVIILRLIILCFFLQYRTTHPVKNAYPLWLTSVICEIWFAFSWLLDQFPKWYPINRET 333
```

```
        233 YIERLSARYEREGEPSQLAGVDFFVSTVDPLKEPPLITANTVLSILAVDYPVDKVSCYVS 292
        334 YLDRLAIRYDRDGEPSQLVPVDVFVSTVDPLKEPPLVTANTVLSILSVDYPVDKVACYVS 393

        293 DDGAAMLSFESLVETAEFARKWVPFCKKFSIEPRAPEFYFSQKIDYLKDKVQPSFVKERR 352
        394 DDGSAMLTFESLSETAEFAKKWVPFCKKFNIEPRAPEFYFAQKIDYLKDKIQPSFVKERR 453

        353 AMKRDYEEYKVRVNALVAKAQKTPEEGWTMQDGTPWPGNNTRDHP-------GHDSGLPW 405
        454 AMKREYEEFKVRINALVAKAQKIPEEGWTMQDGTPWPGNNTRDHPGMIQVFLGHSGGL-- 511

        406 EILGARDIEGNELPRLVYVSREKRPGYQHHKKAGAENALVRVSAVLTNAPYILNVDCDHY 465
        512 ------DTDGNELPRLIYVSREKRPGFQHHKKAGAMNALIRVSVVLTNGAYLLNVDCDHY 565

        466 VNNSKAVREAMCILMDPQVGRDVCYVQFPQRFDGIDKSDRYANRNVVFFDVNMKGLDGIQ 525
  >     566 FNNSKAIKEAMCFMMDPAIGKKCCYVQFPQRFDGIDLHDRYANRNIVFFDINMKGLDGIQ 625

  /     526 GPVYVGTGCVFNRQALYGYGPPSMPSLRKRKDSSSCFSCCCPSKKKPAQDPAEVY---RD 582
  >     626 GPVYVGTGCCFNRQALYGYDP---VLTEEDLEPNIIVKSCCGSRKKGKSSKKYNYEKRRG 682

  /     583 AKREDLNAAIFNLTEID----NYDEHERSMLISQLSFEKTFGLSSVFIESTLMENGGVPE 638
  >     683 INRSDSNAPLFNMEDIDEGFEGYDD-ERSILMSQRSVEKRFGQSPVFIAATFMEQGGIPP 741

  /     639 SANSPPFIKEAIQVIGCGYEEKTEWGKQIGWIYGSVTEDILSGFKMHCRGWRSIYCMPVR 698
  >     742 TTNPATLLKEAIHVISCGYEDKTEWGKEIGWIYGSVTEDILTGFKMHARGWISIYCNPPR 801

  y     699 PAFKGSAPINLSDRLHQVLRWALGSVEIFFSRHCPLWYGFGGGRLKWLQRLAYINTIVYP 758
  b     802 PAFKGSAPINLSDRLNQVLRWALGSIEILLSRHCPIWYGY-HGRLRLLERIAYINTIVYP 860

  y     759 FTSLPLIAYCTIPAVCLLTGKFIIPTLSNLASMLFLGLFISIIVTAVLELRWSGVSIEDL 818
  b     861 ITSIPLIAYCILPAFCLITDRFIIPEISNYASIWFILLFISIAVTGILELRWSGVSIEDW 920

  y     819 WRNEQFWVIGGVSAHLFAVFQGFLKMLAGIDTNFTVTAKAA-EDAEFGELYMVKWTTLLI 877
  b     921 WRNEQFWVIGGTSAHLFAVFQGLLKVLAGIDTNFTVTSKATDEDGDFAELYIFKWTALLI 980

  y     878 PPTTLLIINMSG-CAGFSDALNKGYEAWGPLFGKVFFAFWVILHLYPFLKGLMGRQNLTP 936
  b     981 PPTTVLLVNLIGIVAGVSYAVNSGYQSWGPLFGKLFFALWVIAHLYPFLKGLLGRQNRTP 1040

  y     937 TIVVLWSVLLASVFSLVWVKINPFVNKVDN 966
  b    1041 TIVIVWSVLLASIFSLLWVRINPFVDANPN 1070

earch completed: October 29, 2003, 18:33:06
ob time : 59.0213 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 29, 2003, 18:27:33 ; Search time 28.3752 Seconds
(without alignments)
1458.314 Million cell updates/sec

Title: US-09-980-043-2
Perfect score: 5240
Sequence: 1 MMESGAPICHTCGEQVGHEA.....PPVNRV.NILVAETCISIDC 974

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4524	86.3	974	3	US-08-960-048-6	Sequence 6, Appli
2	4524	86.3	974	4	US-09-838-586-6	Sequence 6, Appli
3	3347	63.9	1081	4	US-09-221-013A-6	Sequence 6, Appli
4	3343	63.8	1081	4	US-09-221-013A-12	Sequence 12, Appl
5	3329	63.5	1065	4	US-09-221-013A-10	Sequence 10, Appl
6	3242	61.9	1084	4	US-09-221-013A-8	Sequence 8, Appli
7	3240.5	61.8	881	3	US-08-960-048-8	Sequence 8, Appli
8	3240.5	61.8	881	4	US-09-838-586-8	Sequence 8, Appli
9	2733	52.2	685	3	US-08-960-048-7	Sequence 7, Appli
10	2733	52.2	685	4	US-09-838-586-7	Sequence 7, Appli
11	2281.5	43.5	629	4	US-09-221-013A-2	Sequence 2, Appli
12	1396.5	26.7	547	4	US-09-221-013A-14	Sequence 14, Appl
13	228.5	4.4	693	3	US-08-960-048-11	Sequence 11, Appl
14	228.5	4.4	693	4	US-09-838-586-11	Sequence 11, Appl
15	223	4.3	3031	1	US-07-689-008-2	Sequence 2, Appli
16	217	4.1	756	3	US-08-960-048-10	Sequence 10, Appl
17	217	4.1	756	4	US-09-147-236-3	Sequence 3, Appli
18	217	4.1	756	4	US-09-522-474-3	Sequence 3, Appli
19	217	4.1	756	4	US-09-838-586-10	Sequence 10, Appl
20	207.5	4.0	861	3	US-08-960-048-12	Sequence 12, Appl
21	207.5	4.0	861	4	US-09-838-586-12	Sequence 12, Appl
22	174.5	3.3	723	3	US-08-960-048-9	Sequence 9, Appli
23	174.5	3.3	723	4	US-08-541-939-2	Sequence 2, Appli
24	174.5	3.3	723	4	US-09-838-586-9	Sequence 9, Appli
25	174.5	3.3	723	5	PCT-US91-01726-4	Sequence 4, Appli
26	127.5	2.4	1177	4	US-09-795-927-2	Sequence 2, Appli
27	113.5	2.2	937	1	US-08-253-155A-31	Sequence 31, Appl
28	113	2.2	501	4	US-09-328-352-5227	Sequence 5227, Ap
29	111	2.1	709	1	US-08-444-005-17	Sequence 17, Appl
30	108.5	2.1	3665	2	US-08-222-617A-13	Sequence 13, Appl
31	108.5	2.1	3712	2	US-08-222-617A-25	Sequence 25, Appl
32	108	2.1	1198	4	US-09-284-768A-10	Sequence 10, Appl
33	107.5	2.1	3712	2	US-08-222-617A-4	Sequence 4, Appli
34	106	2.0	943	3	US-08-476-515A-12	Sequence 12, Appl
35	106	2.0	1065	3	US-09-412-545-2	Sequence 2, Appli
36	106	2.0	4654	3	US-08-476-515A-84	Sequence 84, Appl
37	103.5	2.0	357	1	US-08-119-773-4	Sequence 4, Appli
38	102.5	2.0	490	4	US-09-134-001C-4019	Sequence 4019, Ap
39	101	1.9	452	4	US-09-134-001C-3935	Sequence 3935, Ap
40	101	1.9	545	4	US-09-328-352-6221	Sequence 6221, Ap
41	101	1.9	904	4	US-09-252-991A-19257	Sequence 19257, A
42	100.5	1.9	1172	4	US-09-328-352-6071	Sequence 6071, Ap
43	98	1.9	957	4	US-09-252-991A-24683	Sequence 24683, A
44	97.5	1.9	808	4	US-09-134-001C-3105	Sequence 3105, Ap
45	96.5	1.8	1416	1	US-08-061-465-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-08-960-048-6
; Sequence 6, Application US/08960048C
; Patent No. 6271443
; GENERAL INFORMATION:
; APPLICANT: Stalker, D. et al.
; TITLE OF INVENTION: Plant Cellulose Synthase and Promoter
; TITLE OF INVENTION: Sequences
; FILE REFERENCE: 15621/01/US
; CURRENT APPLICATION NUMBER: US/08/960,048C
; CURRENT FILING DATE: 1997-10-29
; PRIOR APPLICATION NUMBER: 60/029,987
; PRIOR FILING DATE: 1996-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 974
; TYPE: PRT
; ORGANISM: Gossypim hirsutum
US-08-960-048-6

Query Match 86.3%; Score 4524; DB 3; Length 974;
Best Local Similarity 85.5%; Pred. No. 0;
Matches 840; Conservative 70; Mismatches 60; Indels 12; Gaps 7;

```
Qy          1 MMESGAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEINEGRKVCLRCGSPYDE 60
              |||||  |:|||||| || : || |||||||::|:||||||::: |||| ||||||||||
Db          1 MMESGVPVCHTCGEHVGLNVNGEPFVACHECNFPICKSCFEYDLKEGRKACLRCGSPYDE 60

Qy         61 NLLDDVEKKGSGNQSTMASHLNDSQDVGIHARHISSVSTVDSEMNDEYGNPIWKNRVKSC 120
              ||||||| | :|:|||||:||| |||||||||||||||:|||| :: || ||||||:|
Db         61 NLLDDVE-KATGDQSTMAAHLNKSQDVGIHARHISSVSTLDSEMAEDNGNSIWKNRVESW 119

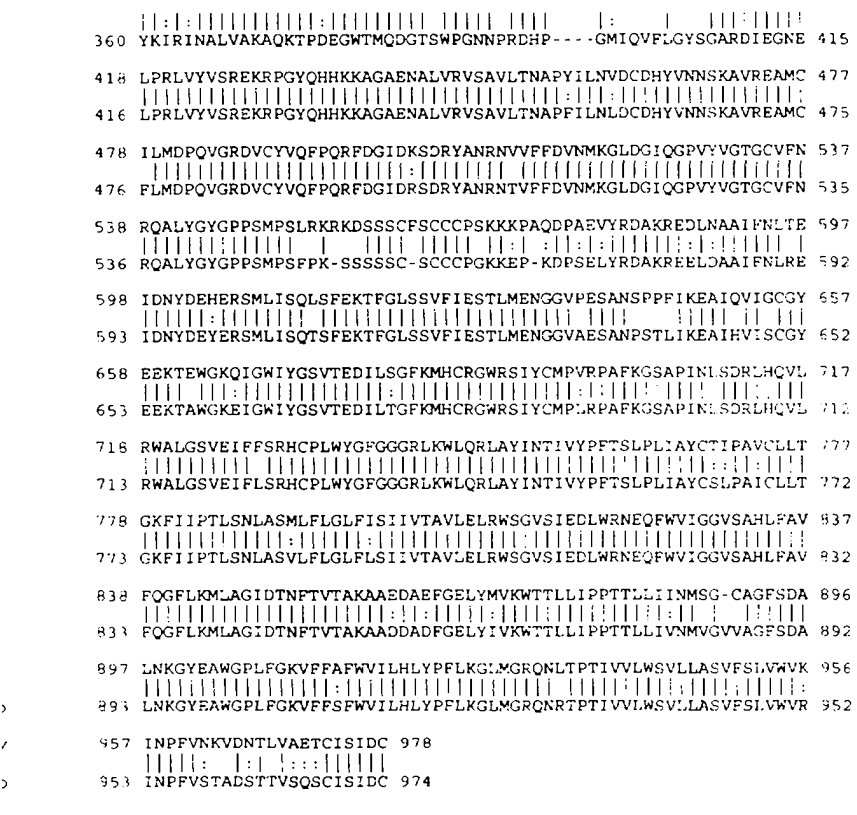
Qy        121 KDKENKKKKRSPKAETEPAQVPTEQQMEEKPSAEASEPLSIVYPIPRNKLTPYRAVIIMR 180
              |:|:||||| :          ::| |||||:|| :||:||| : |||:::| ||| ||||
Db        120 KEKKNKKKKPATTKVEREAEIPPEQQMEDKPAPDASQPLSTIIPIPKSRLAPYRTVIIMR 179

Qy        181 LVILGLFFHFRITNPVDSAFGLWLTSVICEIWFAFSWVLDQFPKWNPVNRETYIERLSAR 240
              |:|||||||:|:|||||| ||||||||||||||||||||||||||| ||||||||:|||||
Db        180 LIILGLFFHYRVTNPVDSAFGLWLTSVICEIWFAFSWVLDQFPKWYPVNRETYIDRLSAR 239

Qy        241 YEREGEPSQLAGVDFFVSTVDPLKEPPLITANTVLSILAVDYPVDKVSCYVSDDGAAMLS 300
              ||||||| :|| |||||||||||||||||||||||||||:|||||||||:||||||||:
Db        240 YEREGEPDELAAVDFFVSTVDPLKEPPLITANTVLSILALDYPVDKVSCYISDDGAAMLT 299

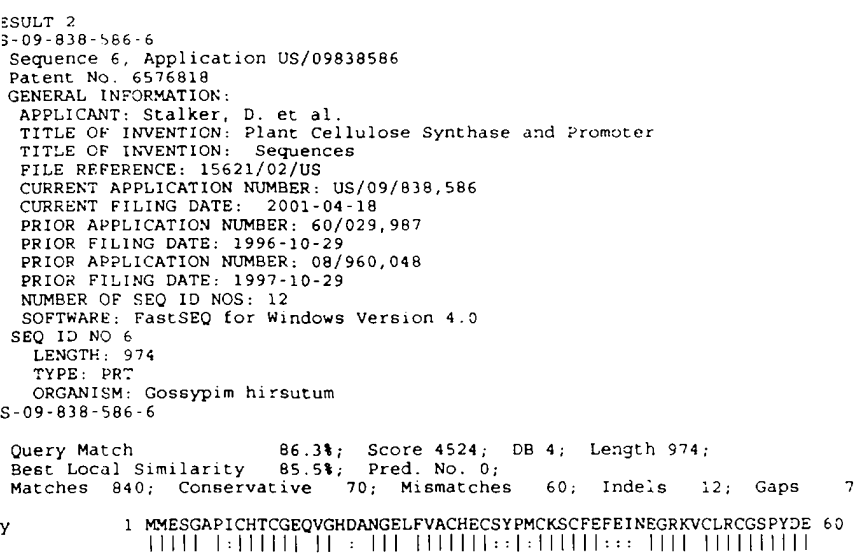
Qy        301 FESLVETAEFARKWVPFCKKFSIEPRAPEFYFSQKIDYLKDKVQPSFVKERRAMKRDYEE 360
              |||:|||:||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        300 FESLVETADFARKWVPFCKKFSIEPRAPEFYFSQKIDYLKDKVQPSFVKERRAMKRDYEE 359

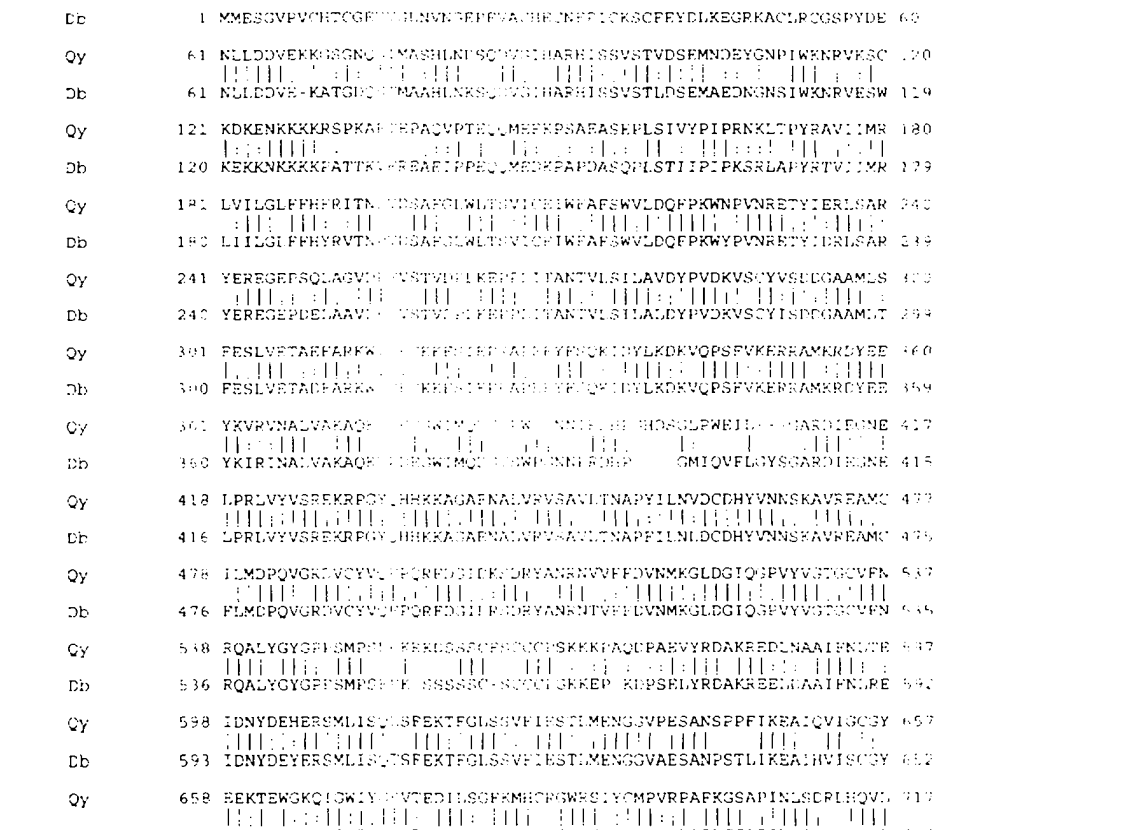
Qy        361 YKVRVNALVAKAQKTPEEGWTMQDGTPWPGNNTRDHPGHDSGLPWEIL---GARDIEGNE 417
```



ESULT 2
3-09-838-586-6
Sequence 6, Application US/09838586
Patent No. 6576818
GENERAL INFORMATION:
APPLICANT: Stalker, D. et al.
TITLE OF INVENTION: Plant Cellulose Synthase and Promoter
TITLE OF INVENTION: Sequences
FILE REFERENCE: 15621/02/US
CURRENT APPLICATION NUMBER: US/09/838,586
CURRENT FILING DATE: 2001-04-18
PRIOR APPLICATION NUMBER: 60/029,987
PRIOR FILING DATE: 1996-10-29
PRIOR APPLICATION NUMBER: 08/960,048
PRIOR FILING DATE: 1997-10-29
NUMBER OF SEQ ID NOS: 12
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 6
LENGTH: 974
TYPE: PRT
ORGANISM: Gossypim hirsutum
S-09-838-586-6

Query Match 86.3%; Score 4524; DB 4; Length 974;
Best Local Similarity 85.5%; Pred. No. 0;
Matches 840; Conservative 70; Mismatches 60; Indels 12; Gaps 7;





RESULT 3
US-09-221-013A-6
; Sequence 6, Application US/09221013A
; Patent No. 6495740
; GENERAL INFORMATION:
; APPLICANT: Arioli, Antonio

; APPLICANT: Williamson, Richard E.
; APPLICANT: Betzner, Andreas S.
; APPLICANT: Peng, Liangcai
; TITLE OF INVENTION: Manipulation of cellulose and/or Beta-1,4-glucan
; FILE REFERENCE: 96-98
; CURRENT APPLICATION NUMBER: US/09/221,013A
; CURRENT FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: PCT/AU97/00402
; PRIOR FILING DATE: 1997-06-24
; PRIOR APPLICATION NUMBER: AU PO0699
; PRIOR FILING DATE: 1996-06-27
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 1081
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-221 013A-6

Query Match 63.9%; Score 3347; DB 4; Length 1081;
Best Local Similarity 61.3%; Pred. No. 3.2e-314;
Matches 644; Conservative 122; Mismatches 176; Indels 108; Gaps 18;

```
Qy          8 ICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEINEGRKVCLRC--------GSP-- 57
              |  ||: ||    |::|||::||::|:|: |:|:|  :| : | :|        |||
Db         38 ICQICGDDVGLAETGDVFVACNECAFPVCRPCYEYERKDGTQCCPQCKTRFRRHRGSPRV 97

Qy         58 -- YDENLLDDVEKKGSGNQ-STMASHLNDSQDVGIHARHIS ------------SVST 99
                ||: :||:| : :  | :  | :    ::     :|| |             : |
Db         98 EGDEDEDDVDDIENEFNYAQGANKARHQRHGEEFSSSSRHESQPIPLLTHGHTVSGEIRT 157

Qy        100 VDSE  ---------------MNDEY --    --  ------ - ----GNPIWKNRVKSC 120
                |::                ::  |                    ||  || ||:
Db        158 PDTQSVRTTSGPLGPSDRNAISSPYIDPRQPVPVRIVDPSKDLNSYGLGNVDWKERVEGW 217

Qy        121 KDKENK-------KKKRSPKAETEPAQVPTEQQMEEKPSAEASE-PLSIVYPIPRNKLTP 172
              | |: |       |       | |     |    ||   |: :  |:| | ||| ::|||
Db        218 KLKQEKNMLQMTGKYHEGKGGEIE----GTGSNGEELQMADDTRLPMSRVVPIPSSRLTP 273

Qy        173 YRAVIIMRLVILGLFFHFRITNPVDSAFGLWLTSVICEIWFAFSWVLDQFPKWNPVNRET 232
              || |||:||:||  |  :| |:|| :|: |||:|||||.|||||:|||:||| |:||||
Db        274 YRVVIILRLIILCFFLQYRTTHPVKNAYPLWLTSVICEIWFAFSWLLDQFPKWYPINRET 333

Qy        233 YIERLSARYEREGEPSQLAGVDFFVSTVDPLKEPPLITANTVLSILAVDYPVDKVSCYVS 292
              |:||: ||:|:|||||  || ||||||||||:|:|:|:|||:||:|||||||:||||
Db        334 YLDRLAIRYDRDGEPSQLVPVDVFVSTVDPLKEPPLVTANTVLSILSVDYPVDKVACYVS 393

Qy        293 DDGAAMLSFESLVETAEFARKWVPFCKKFNIEPRAPEFYFAQKIDYLKDKVQPSFVKERR 352
              :|:|||:|||| ||||||:|||| :|:|:|| ||||    | |||||:||||||||||||
Db        394 DDGSAMLTFESLSETAEFAKKWVPFCKKFNIEPRAPEFYFAQKIDYLKDKIQPSFVKERR 453

Qy        353 AMKRDYEEYKVRVNALVAKAQKTPEEGWTMQDGTPWPGNNTRDHP-------GHDSGLPW 405
              |:||:|||:|||:|||||||| |||| ||||:|:|||:|:|||        || ||
Db        454 AMKREYEEFKVRINALVAKAQKIPEEGWTMQDGTPWPGNNTRDHPGMIQVFLGHSGGL-- 511

Qy        406 EILGARDIEGNELPRLVYVSREKRPGYQHHKKAGAENALVRVSAVLTNAPYILNVDCDHY 465
                  : :||||||:|||||:||.| ||||| ||| :||||||  |:||||'|||
Db        512 - -- -DTDGNELPRLIYVSREKRPGFQHHKKAGAMNALIRVSAVLTNGAYLLNVDCDHY 565

Qy        466 VNNSKAVREAMCILMDPQVGRDVCYVQFPQRFDGIDKSDRYANRNVVFFDVNMKGLDGIQ 525
              |||||::|||| :||| :|:  |||||||:|||||  ||  |||:|||:|||||||||
Db        566 FNNSKAIKEAMCFMMDPAIGKKCCYVQFPQRFDGIDLHDRYANRNIVFFDINMKGLDGIQ 625

Qy        526 GPVYVGTGCVFNRQALYGYGPPSMPSLRKRKDSSSCFSCCCPSKKKPAQDPAEVY---RD 582
              | ||||||| ||||:|||      .  : :      ||  :||        |   |
Db        626 GPVYVGTGCCFNRQALYGYDP---VLTEEDLEFNIIVKSCWISRKKGKSSKKYNYEKRRG 682

Qy        583 AKREDLNAAIFNLTEID--- NYDEHERSMLISQLSFEKTFGLSSVPIESTLMENGGVPE 638
              .| | || :||: :||     ||: |||:|: | | || ||  ||: :| || ||:
Db        683 INRSDSNAPLFNMEDIDEGFEGYDD-ERSILMSQRSVEKRFGQSPVFIAATFMEQGGIPP 741

Qy        639 SANSPPFIKEAIQVIGCGYEEKTEWGKQIGWIYGSVTEDILTGFKMHCRGWRSIYCMPVR 698
              : |    :|||| || ||||:|||||.:|||| ||| :|:|:|||| |:| ||'| | |
Db        742 TTNPATLLKEAIHVISCGYEDKTEWGKEIGWIYGSVTEDILTGFKMHARGWISIYCNPPR 801

Qy        699 PAFKGSAPINLSDRLHQVLRWALGSVEIFFSRHCPLWYGFGGGRLKWLQRLAYINTIVYP 758
              |||||||||||||||:|||||||:||   |||||:|||:  |||: |:|:|||||||||
Db        802 PAFKGSAPINLSDRLNQVLRWALGSIEILLSRHCPIWYGY-HGRLRLLERIAYINTIVYP 860

Qy        759 FTSLPLIAYCTIPAVCLLTGKFIIPTLSNLASMLFLGLFISIIVTAVLELRWSGVSIEDL 818
              ||:|||||| :|| ||:|  :|||| :|| ||: |: ||||| || :|||||||||||
Db        861 ITSIPLIAYCILPAFCLITDRFIIPEISNYASIWFILLFISIAVTGILELRWSGVSIEDW 920

Qy        819 WRNEQFWVIGGVSAHLFAVFQGFLKMLAGIDTNFTVTAKAA-EDAEFGELYMVKWTTLLI 877
              |||||||||| |||||||||| ||:|||||||||||:||  || :| |||: ||| |||
Db        921 WRNEQFWVIGGTSAHLFAVFQGLLKVLAGIDTNFTVTSKATDEDGDFAELYIFKWTALLI 980

Qy        878 PPTTLLIINMSG-CAGFSDALNKGYEAWGPLFGKVFFAFWVILHLYPFLKGLMGRQNLTP 936
              ||||:|::|: |  || | |:| ||::|||||||:||| ||| |||||||||:||| ||
Db        981 PPTTVLLVNLIGIVAGVSYAVNSGYQSWGPLFGKLFFALWVIAHLYPFLKGLLGRQNRTP 1040

Qy        937 TIVVLWSVLLASVFSLVWVKINPFVNKVDN 966
              |||::|||||||:|||:||:|||||:   |
Db       1041 TIVIVWSVLLASIFSLLWVRINPFVDANPN 1070
```

RESULT 4
US-09-221-013A-12
; Sequence 12, Application US/09221013A
; Patent No. 6495740
; GENERAL INFORMATION:
; APPLICANT: Arioli, Antonio
; APPLICANT: Williamson, Richard E.
; APPLICANT: Betzner, Andreas S.
; APPLICANT: Peng, Liangcai
; TITLE OF INVENTION: Manipulation of cellulose and/or Beta-1,4-glucan
; FILE REFERENCE: 96-98
; CURRENT APPLICATION NUMBER: US/09/221,013A
; CURRENT FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: PCT/AU97/00402
; PRIOR FILING DATE: 1997-06-24
; PRIOR APPLICATION NUMBER: AU PO0699
; PRIOR FILING DATE: 1996-06-27
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 1081
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-221-013A-12

Query Match 63.8%; Score 3343; DB 4; Length 1081;
Best Local Similarity 61.2%; Pred. No. 7.8e-314;
Matches 643; Conservative 122; Mismatches 177; Indels 108; Gaps 18;

```
Qy          8 ICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEINEGRKVCLRC--------GSP-- 57
              ||  ||: ||    |::| ||:||::|:|: |:|:|  :| : | :|        |||
Db         38 ICQICGDDVGLAETGDVFVACNECAFPVCRPCYEYERKDGTQCCPQCKTRFRRHRGSPRV 97

Qy         58 ---YDENLLDDVEKKGSGNQ-STMASHLNDSQDVGIHARHIS-------------SVST 99
                  ||: :||:| : :  | :  | |    ::    :|| |             : |
Db         98 EGDEDEDDVDDIENEFNYAQGANKARHQRHGEEFSSSSRHESQPIPLLTHGHTVSGEIRT 157

Qy        100 VDSE----------------MNDEY----------------------GNPIWKNRVKSC 120
                |::                ::  |                    ||  || ||:
Db        158 PDTQSVRTTSGPLGPSDRNAISSPYIDPRQPVPVRIVDPSKDLNSYGLGNVDWKERVEGW 217

Qy        121 KDKENK-------KKKRSPKAETEPAQVPTEQQMEEKPSAEASE-PLSIVYPIPRNKLTP 172
              | |: |       |       | |     |    ||   |: :  |:| | ||| ::|||
Db        218 KLKQEKNMLQMTGKYHEGKGGEIE----GTGSNGEELQMADDTRLPMSRVVPIPSSRLTP 273

Qy        173 YRAVIIMRLVILGLFFHFRITNPVDSAFGLWLTSVICEIWFAFSWVLDQFPKWNPVNRET 232
              || |||:||:||  |  :| |:|| :|: |||:|||||||||:|||:||| |:||||
Db        274 YRVVIILRLIILCFFLQYRTTHPVKNAYPLWLTSVICEIWFAFSWLLDQFPKWYPINRET 333

Qy        233 YIERLSARYEREGEPSQLAGVDFFVSTVDPLKEPPLITANTVLSILAVDYPVDKVSCYVS 292
```

```
o      334 YLDRLAIRYDRDGEPSQLVPVDVFVSTVDPLKEPPLVTANTVLSILSVDYPVDKVACYVS 393

y      293 DDGAAMLSFESLVETAEFARKWVPFCKKFSIEPRAPEFYFSQKIDYLKDKVQPSFVKERR 352
o      394 DDGSAMLTFESLSETAEFAKKWVPFCKKFNIEPRAPEFYFAQKIDYLKDKIQPSFVKERR 453

y      353 AMKRDYEEYKVRVNALVAKAQKTPEEGWTMQDGTPWPGNNTRDHP-------GHDSGLPW 405
o      454 AMKREYEEFKVRINALVAKAQKIPEEGWTMQDGTPWPGNNTRDHPGMIQVFLGHSGGL-- 511

y      406 EILGARDIEGNELPRLVYVSREKRPGYQHHKKAGAENALVRVSAVLTNAPYILNVDCDHY 465
o      512 ------DTDGNELPRLIYVSREKRPGFQHHKKAGAMNALIRVSVVLTNGAYLLNVDCDHY 565

y      466 VNNSKAVREAMCILMDPQVGRDVCYVQFPQRFDGIDKSDRYANRNVVFFDVNMKGLDGIQ 525
o      566 FNNSKAIKEAMCFMMDPAIGKKCCYVQFPQRFDGIDLHDRYANRNIVFFDINMKGLDGIQ 625

y      526 GPVYVGTGCVFNRQALYGYGPPSMPSLRKRKDSSSCFSCCCPSKKKPAQDPAEVY---RD 582
o      626 GPVYVGTGCCFNRQALYGYDP---VLTEEDLEPNIIVKSCCGSRKKGKSSKKYNYEKRRG 682

y      583 AKREDLNAAIFNLTEID----NYDEHERSMLISQLSFEKTFGLSSVFIESTLMENGGVPE 638
o      683 INRSDSNAPLFNMEDIDEGFEGYDD-ERSILMSQRSVEKRFGQSPVFIAATFMEQGGIPP 741

y      639 SANSPPFIKEAIQVIGCGYEEKTEWGKQIGWIYGSVTEDILSGFKMHCRGWRSIYCMPVR 698
o      742 TTNPATLLKEAIHVISCGYEDKTEWGKEIGWIYGSVTEDILTGFKMHARGWISIYCNPPR 801

y      699 PAFKGSAPINLSDRLHQVLRWALGSVEIFFSRHCPLWYGFGGGRLKWLQRLAYINTIVYP 758
b      802 PAFKGSAPINLSDRLNQVLRWALGSIEILLSRHCPIWYGY-HGRLRLLERIAYINTIVYP 860

y      759 FTSLPLIAYCTIPAVCLLTGKFIIPTLSNLASMLFLGLFISIIVTAVLELRWSGVSIEDL 818
b      861 ITSIPLIAYCILPAFCLITDRFIIPEISNYASIWFILLFISIAVTGILELRWSGVSIEDW 920

y      819 WRNEQFWVIGGVSAHLFAVFQGFLKMLAGIDTNFTVTAKAA-EDAEFGELYMVKWTTLLI 877
b      921 WRNEQFWVIGGTSAHLFAVFQGLLKVLAGIDTNFTVTSKATDEDGDFAELYIFKWTALLI 980

y      878 PPTTLLIINMSG-CAGFSDALNKGYEAWGPLFGKVFFAFWVILHLYPFLKGLMGRQNLTP 936
b      981 PPTTVLLVNLIGIVAGVSYAVNSGYQSWGPLFGKLFFALWVIAHLYPFLKGLLGRQNRTP 1040

y      937 TIVVLWSVLLASVFSLVWVKINPFVNKVDN 966
b     1041 TIVIVWSVLLASIFSLLWVRINPFVDANPN 1070
```

ESULT 5
S-09-221-013A-10
 Sequence 10, Application US/09221013A
 Patent No. 6495740
 GENERAL INFORMATION:
  APPLICANT: Arioli, Antonio
  APPLICANT: Williamson, Richard E.
  APPLICANT: Betzner, Andreas S.
  APPLICANT: Peng, Liangcai
  TITLE OF INVENTION: Manipulation of cellulose and/or Beta-1,4-glucan
  FILE REFERENCE: 96-98
  CURRENT APPLICATION NUMBER: US/09/221,013A
  CURRENT FILING DATE: 1998-12-23
  PRIOR APPLICATION NUMBER: PCT/AU97/00402
  PRIOR FILING DATE: 1997-06-24
  PRIOR APPLICATION NUMBER: AU PO0699
  PRIOR FILING DATE: 1996-06-27
  NUMBER OF SEQ ID NOS: 37
  SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO 10
   LENGTH: 1065
;   TYPE: PRT
;   ORGANISM: Arabidopsis thaliana
US-09-221-013A-10

  Query Match   63.5%; Score 3329; DB 4; Length 1065;
  Best Local Similarity   60.4%; Pred. No. 1.7e-312;
  Matches 652; Conservative 123; Mismatches 187; Indels 116; Gaps   21;

```
Qy          2 MES---  GAPI--    --CHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEINEGRK 49
Db          1 MESEGETAGKPMKNIVPQTCQICSDNVGKTVDGDRFVACDICSFPVCRPCYEYERKDGNQ 60

Qy         50 VCLRC-- -----GSP-----YDENLLDDV--------------------------EKK 69
Db         61 SCPQCKTRYKRLKGSPAIPGDKDEDGLADEGTVEFNYPQKEKISERMLGWHLTRGKGEEM 120

Qy         70 GSGNQSTMASH--    LNDSQDV    - GIHARHISSVSTV--------DSEMNDE- -- 107
Db        121 GEPQYDKEVSHNHLPRLTSRQDTSGEFSAASPERLSVSSTIAGGKRLPYSSDVNQSPNRR 180

Qy        108 ---    YGNIIWKERVESWKEKKNKKKKPPKAEKEAQVPTEQQM-----EEKPSAEA 155
Db        181 IVDPVGLGNVAWKERVDGWKMKQEKNTGPVSTQAASERGGVDIDASTDILADEALLNDEA 240

Qy        156 SEPLSIVYPIPRNKLTPYRAVIIMRLVILGLFFHFRITNPVDSAFGLWLTSVICEIWFAF 215
Db        241 RQLLSRKVSIPSSRINPYRMVIMLRLVILCLFLHYRITNPVPNAFALWLVSVICEIWFAL 300

Qy        216 SWVLDQFPKWNPVNRETYIERLSARYEREGEPSQLAGVDFFVSTVDPLKEPPLITANTVL 275
Db        301 SWILDQFPKWFPVNRETYLDRLALRYDREGEPSQLAAVDIFVSTVDPLKEPPLVTANTVL 360

Qy        276 SILAVDYPVDKVSCYVSDDGAAMLSFESLVETAEFARKWVPFCKKFSIEPRAPEFYFSQK 335
Db        361 SILAVDYPVDKVSCYVSDDGAAMLSFESLAETSEFARKWVPFCKKYSIEPRAPEWYFAAK 420

Qy        336 IDYLKDKVQPSFVKERRAMKRDYEEYKVRVNALVAKAQKTPEEGWTMQDGTPWPGNNTRD 395
Db        421 IDYLKDKVQTSFVKDRRAMKREYEEFKIRINALVSKALKCPEEGWVMQDGTPWPGNNTGD 480

Qy        396 HP------GHDSGLPWEILGARDIEGNELPRLVYVSREKRPGYQHHKKAGAENALVRVS 448
Db        481 HPGMIQVFLGQNGGL   ---  --DAEGNELPRLVYVSREKRPGFQHHKKAGAMNALVRVS 532

Qy        449 AVLTNAPYILNVDCDHYVNNSKAVREAMCILMDPQVGRDVCYVQFPQRFDGIDKSDRYAN 508
Db        533 AVLTNGPFILNLDCDHYINNSKALREAMCFLMDPNLGKQVCYVQFPQRFDGIDKNDRYAN 592

Qy        509 RNVVFFDVNMKGLDGIQGPVYVGTGCVFNRQALYGYGPPSMPSLRKRKDSSSCFSCCCPS 568
Db        593 RNTVFFDINLRGLDGIQGPVYVGTGCVFNRTALYGYEPPIKV---KHKKPSLLSKLCGGS 649

Qy        569 KKK--PAQDPAEVYRDAKREDLNAAIFNLTEID    ---NYDEHERSMLISQLSFEKTFG 620
Db        650 RKKNSKAKKESDKKKSGRHTDSTVPVFNLDDIEEGVEGAGFDD-EKALLMSQMSLEKRFG 708

Qy        621 LSSVFIESTLMENGGVPESANSPPFIKEAIQVIGCGYEEKTEWGKQIGWIYGSVTEDILS 680
Db        709 QSAVFVASTLMENGGVPPSATPENFLKEAIHVISCGYEDKSDWGMEIGWIYGSVTEDILT 768

Qy        681 GFKMHCRGWRSIYCMPVRPAFKGSAPINLSDRLHQVLRWALGSVEIFFSRHCPLWYGFGG 740
Db        769 GFKMHARGWRSIYCMPKLPAFKGSAPINLSDRLNQVLRWALGSVEILFSRHCPIWYGY-N 827

Qy        741 GRLKWLQRLAYINTIVYPFTSLPLIAYCTIPAVCLLTGKFIIPTLSNLASMLFLGLFISI 800
Db        828 GRLKFLERFAYVNTTIYPITSIPLLMYCTLLAVCLFTNQFIIPQISNIASIWFLSLFLSI 887

Qy        801 IVTAVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQGFLKMLAGIDTNFTVTAKAA- 859
Db        888 FATGILEMRWSGVGIDEWWRNEQFWVIGGVSAHLFAVFQGILKVLAGIDTNFTVTSKASD 947

Qy        860 EDAEFGELYMVKWTTLLIPPTTLLIINMSG-CAGFSDALNKGYEAWGPLFGKVFFAFWVI 918
```



```
Db        948 EDGDFAELYLFKWTTLLIPPTTLLIVNLVGVVAGVSYAINSGYQSWGPLFGKLFFAFWVI 1007

Qy        919 LHLYPFLKGLMGRQNLTPTIVVLWSVLLASVFSLVWVKINPFVNKVDNTLVAETCISIDC 978

Db       1008 VHLYPFLKGLMGRQNRTPTIVVVWSVLLASIFSLLWVRIDPFTSRVTGPDILE--CGINC 1065


RESULT 6
US-09-221-013A-8
; Sequence 8, Application US/09221013A
; Patent No. 6495740
; GENERAL INFORMATION:
;  APPLICANT:  Arioli, Antonio
;  APPLICANT:  Williamson, Richard E.
;  APPLICANT:  Betzner, Andreas S.
;  APPLICANT:  Peng, Liangcai
;  TITLE OF INVENTION:  Manipulation of cellulose and/or Beta-1,4-glucan
;  FILE REFERENCE:  96-98
;  CURRENT APPLICATION NUMBER: US/09/221,013A
;  CURRENT FILING DATE:  1998-12-23
;  PRIOR APPLICATION NUMBER:  PCT/AU97/00402
;  PRIOR FILING DATE:  1997-06-24
;  PRIOR APPLICATION NUMBER:  AU PO0699
;  PRIOR FILING DATE:  1996-06-27
;  NUMBER OF SEQ ID NOS:  37
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 8
;   LENGTH: 1084
;   TYPE: PRT
;   ORGANISM: Arabidopsis thaliana
US-09-221-013A-8

  Query Match           61.9%;  Score 3242;  DB 4;  Length 1084;
  Best Local Similarity  56.9%;  Pred. No. 4.6e-304;
  Matches  615;  Conservative  158;  Mismatches  177;  Indels  130;  Gaps   22;

Qy          1 MMESGAPICHTCGEQVGHDANGELFVACNECSYPVCKSCFEFEINEGRKVCLRC------ 54

Db         31 VQELSGQTCQICGDEIELTVSSELFVACNECAFPVCRPCYEYERREGNQACPQCKTRYKR 90

Qy         55 --GSP------YDENLLDDVEKK -- - -GSGNQSTMACHLN -- --DS----  --- 84

Db         91 IKGSPRVDGDDEEEEDIDDLEYEFDGMGHEHAAEAALSPFLNTGRGLDSAPPGSQIPL 150

Qy         85       QDVGIHA-RH--ISSVST           VDSEMN E-- -  -------YGNP 111

Db        151 LTYGDEDADMYSDRHALIVPPSTGYGNRVYPAPFTDSSAPPQARSMVPGKDIAEYGYGSV 210

Qy        112 IWKNRVKSCKDKENKKKK--- ---- -RSPKAETEPAQVPTEQQMEEKPSAEASEPLSI 161

Db        211 AWKDRMEVWKRRQGEKLQVIKHEGGNNGRGSNDDDE-LDDPDMPMMD-----EGRQPLSR 264

Qy        162 VYPIPRNKLTPYRAVIIMRLVILGLFFHFRITNPVDSAFGLWLTSVICEIWFAFSWVLDQ 221

Db        265 KLPIRSSRINPYRMLILCRLAILGLFFHYRILHPVNDAYGLWLTSVICEIWFAVSWILDQ 324

Qy        222 FPKWNPVNRETYIERLSARYEREGEPSQLAGVDFFVSTVDPLKEPPLITANTVLSILAVD 281

Db        325 FPKWYPIERETYLDRLSLRYEKEGKPSGLAPVDVFVSTVDPLKEPPLITANTVLSILAVD 384

Qy        282 YPVDKVSCYVSDDGAAMLSFESLVETAEFARKWVPFCKKFSIEPRAPEFYFSQKIDYLKD 341

Db        385 YPVDKVACYVSNNGAAMLTFEALSDTADFATKWVPFCKKFNIEPRAPEWYFSQKMDYLKN 444

Qy        342 KVQPSFVKERRAMKRDYEEYKVRVNALVAKAQKTPEEGWTMQDGTPWPGNNTRDHP---- 397

Db        445 KVHPAFVRERRAMKRDYEEFKVKINALVATAQKVPEEGWTMQDGTPWPGNNVRDHPGMIQ 504

Qy        398 ---GHDSGLPWEILGARDIEGNELPRLVYVSREKRPGYQHHKKAGAENALVRVSAVLTNA 454

Db        505 VFLGHS--------GVRDTDGNELPRLVYVSREKRPGFDHHKKAGAMNSLIRVSAVLSNA 556
```

```
Qy        455 PYILNVDCDHYVNNSKAVREAMCILMDPQVGRDVCYVQFPQRFDGIDKSDRYANRNVVFF 514

Db        557 PYLLNVDCDHYINNSKAIRESMCFMMDPQSGKKVCYVQFPQRFDGIDRHDRYSNRNVVFF 616

Qy        515 DVNMKGLDGIQGPVYVGTGCVFNRQALYGYGPPSMPSLRKRKDSSSCFSC-------CCP 567

Db        617 DINMKGLDGIQGPIYVGTGCVFRKQALYGFDAP-----KKKKPPGKTCNCWPKWCCLCCG 671

Qy        568 SKKKPAQDPAEVYRDAKREDLNAAIFNLTEIDNYDEH--------ERSMLISQLSFEKTF 619

Db        672 LRKKSKTKATD-----KKTNTKETSKQIHALENVDEGVIVPVSNVEKRSEATQLKLEKKF 726

Qy        620 GLSSVFIESTLMENGGVPESANSPPFIKEAIQVIGCGYEEKTEWGKQIGWIYGSVTEDIL 679

Db        727 GQSPVFVASAVLQNGGVPRNASPACLLREAIQVISCGYQDKTEWGKEIGWIYGSVTEDIL 786

Qy        680 SGFKMHCRGWRSIYCMPVRPAFKGSAPINLSDRLHQVLRWALGSVEIFFSRHCPLWYGFG 739

Db        787 TGFKMHCHGWRSVYCMPKRAAFKGSAPINLSDRLHQVLRWALGSVEIFLSRHCPIWYGYG 846

Qy        740 GGRLKWLQRLAYINTIVYPFTSLPLIAYCTIPAVCLLTGKFIIPTLSNLASMLFLGLFIS 799

Db        847 GG-LKWLERFSYINSVVYPWTSLPLIVYCSLPAVCLLTGKFIVPEISNYAGILFMLMFIS 905

Qy        800 IIVTAVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQGFLKMLAGIDTNFTVTAKAA 859

Db        906 IAVTGILEMQWGGVGIDDWWRNEQFWVIGGASSHLFALFQGLLKVLAGVNTNFTVTSKAA 965

Qy        860 EDAEFGELYMVKWTTLLIPPTTLLIINMSG-CAGFSDALNKGYEAWGPLFGKVFFAFWVI 918

Db        966 DDGAFSELYIFKWTTLLIPPTTLLIINIIGVIVGVSDAISNGYDSWGPLFGRLFFALWVI 1025

Qy        919 LHLYPFLKGLMGRQNLTPTIVVLWSVLLASVFSLVWVKINPFVNKVDNTLVAETCISIDC 978

Db       1026 VHLYPFLKGMLGKQDKMPTIVVWSILLASILTLLWVRINPFVAK--GGPVLEIC-GLNC 1082


RESULT 7
US-08-960-048-8
; Sequence 8, Application US/08960048C
; Patent No. 6271443
; GENERAL INFORMATION:
;  APPLICANT:  Stalker, D. et al.
;  TITLE OF INVENTION:  Plant Cellulose Synthase and Promoter
;  TITLE OF INVENTION:  Sequences
;  FILE REFERENCE:  15621/01/US
;  CURRENT APPLICATION NUMBER: US/08/960,048C
;  CURRENT FILING DATE:  1997-10-29
;  PRIOR APPLICATION NUMBER:  60/029,987
;  PRIOR FILING DATE:  1996-10-29
;  NUMBER OF SEQ ID NOS:  12
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 8
;   LENGTH: 881
;   TYPE: PRT
;   ORGANISM: Oryzae sativa
US-08-960-048-8

  Query Match           61.8%;  Score 3240.5;  DB 3;  Length 881;
  Best Local Similarity  66.5%;  Pred. No. 4.5e-304;
  Matches  605;  Conservative  108;  Mismatches  128;  Indels   69;  Gaps   13;

Qy        109 GNPIWKNRVKSCKDKENKKKKRSPKAETEPAQVPTEQQMEEKPSA---------- 153

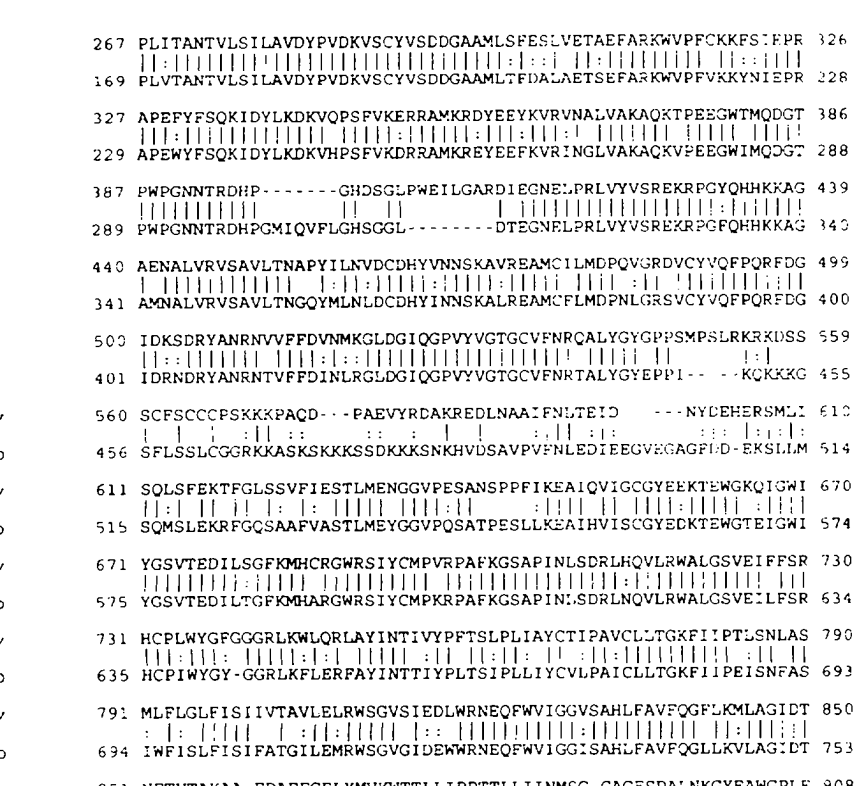
Db          1 GNVAWKERVDGWKLKQDK-----------GAIPMTNGTSIAPSEGRGVGDIDASTDYNN 48

Qy        154 -------EASEPLSIVYPIPRNKLTPYRAVIIMRLVILGLFFHFRITNPVDSAFGLWLTS 206

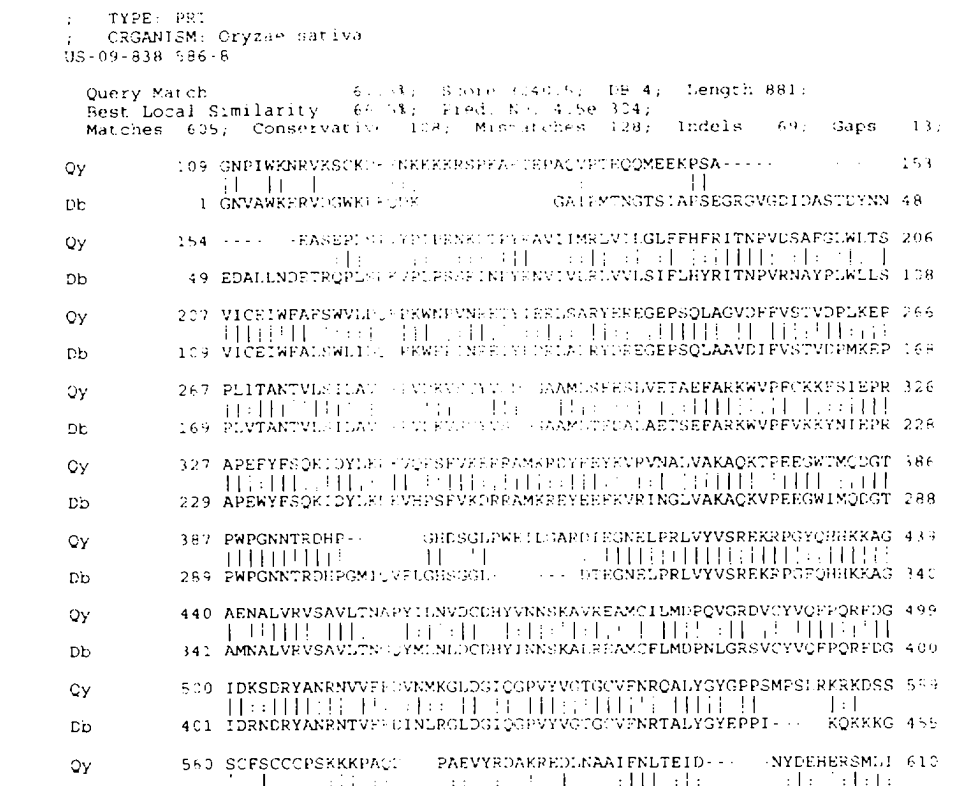
Db         49 EDALLNDETRQPLSRKVPLPSSRINPYRNVIVLRLVVLSIFLHYRITNPVRNAYPLWLLS 108

Qy        207 VICEIWFAFSWVLDQFPKWNPVNRETYIERLSARYEREGEPSQLAGVDFFVSTVDPLKEP 266

Db        109 VICEIWFALSWLIDQFPKWFPINRETYLDRLALRYDREGEPSQLAAVDIFVSTVDPMKEP 168
```



ESULT 8
S-09-838-586-8
Sequence 8, Application US/09838586
Patent No. 6576818
GENERAL INFORMATION:
APPLICANT: Stalker, D. et al.
TITLE OF INVENTION: Plant Cellulose Synthase and Promoter
TITLE OF INVENTION: Sequences
FILE REFERENCE: 15621/02/US
CURRENT APPLICATION NUMBER: US/09/838,586
CURRENT FILING DATE: 2001-04-18
PRIOR APPLICATION NUMBER: 60/029,987
PRIOR FILING DATE: 1996-10-29
PRIOR APPLICATION NUMBER: 08/960,048
PRIOR FILING DATE: 1997-10-29
NUMBER OF SEQ ID NOS: 12
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 8
LENGTH: 881



; TYPE: PRT
; ORGANISM: Oryzae sativa
US-09-838-586-8

Query Match 6...3; Score ...; DB 4; Length 881;
Best Local Similarity 69.7%; Pred. No. 4.5e-324;
Matches 695; Conservative 134; Mismatches 128; Indels 69; Gaps 13;

```
RESULT 9
US-08-960-048-7
; Sequence 7, Application US/08960048C
; Patent No. 6271443
; GENERAL INFORMATION:
;  APPLICANT:  Stalker, D. et al.
;  TITLE OF INVENTION:  Plant Cellulose Synthase and Promoter
;  TITLE OF INVENTION:  Sequences
;  FILE REFERENCE:  15621/01/US
;  CURRENT APPLICATION NUMBER: US/08/960,048C
;  CURRENT FILING DATE:  1997-10-29
;  PRIOR APPLICATION NUMBER:  60/029,987
;  PRIOR FILING DATE:  1996-10-29
;  NUMBER OF SEQ ID NOS:  12
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 7
;   LENGTH: 685
;   TYPE: PRT
;   ORGANISM: Gossypium hirsutum
US-08-960-048-7

  Query Match             52.2%;  Score 2733;  DB 3;  Length 685;
  Best Local Similarity   75.2%;  Pred. No. 3.8e-255;
  Matches  509;  Conservative   66;  Mismatches   74;  Indels   28;  Gaps    8;

Qy        311 ARKWVPFCKKFSIEPRAPEFYFSQKIDYLKDKVQPSFVKERRAMKRDYEEYKVRVNALVA 370
Db          1 ARRWVPFCKKHNVEPRAPEFYFNEKIDYLKDKVHPSFVKERRAMKREYEEFKVRINALVA 60

Qy        371 KAQKTPEEGWTMQDGTPWPGNNTRDHPGHDSGLPWEILGARDIEGNELPRLVYVSREKRP 430
Db         61 KAQKKPEEGWVMQDGTPWPGNNTRDHPGMIQ-VYLGSAGALDVDGKELPRLVYVSREKRP 119

Qy        431 GYQHHKKAGAENALVRVSAVLTNAPYILNVDCDHYVNNSKAVREAMCILMDPQVGRDVCY 490
Db        120 GYQHHKKAGAENALVRVSAVLTNAPFILNLDCDHYINNSKAMREAMCFLMDPQFGKKLCY 179

Qy        491 VQFPQRFDGIDKSDRYANRNVVFFDVNMKGLDGIQGPVYVGTGCVFNRQALYGYGPPSMP 550
Db        180 VQFPQRFDGIDRHDRYANRNVVFFDINMLGLDGLQGPVYVGTGCVFNRQALYGYDPP--V 237

Qy        551 SLRKRKDSSSCF----SCCCPSKKKPAQDPAE--------VYRDAKR-------EDLNAA 591
Db        238 SEKRPKMTCDCWPSWCCCCCCGGSRKKSKKKGEKKGLLGGLLYGKKKMMGKNYVKKGSAP 297

Qy        592 IFNLTEI----DNYDEHERSMLISQLSFEKTFGLSSVFIESTLMENGGVPESANSPPFIK 647
Db        298 VFDLEEIEEGLEGYEELEKSTLMSQKNFEKRFGQSPVFIASTLMENGGLPEGTNSTSLIK 357

Qy        648 EAIQVIGCGYEEKTEWGKQIGWIYGSVTEDILSGFKMHCRGWRSIYCMPVRPAFKGSAPI 707
Db        358 EAIHVISCGYEEKTEWGKEIGWIYGSVTEDILTGFKMHCRGWKSVYCVPKRPAFKGSAPI 417

Qy        708 NLSDRLHQVLRWALGSVEIFFSRHCPLWYGFGGGRLKWLQRLAYINTIVYPFTSLPLIAY 767
Db        418 NLSDRLHQVLRWALGSVEIFLSRHCPLWYGY-GGKLKWLERLAYINTIVYPFTSIPLLAY 476

Qy        768 CTIPAVCLLTGKFIIPTLSNLASMLFLGLFISIIVTAVLELRWSGVSIEDLWRNEQFWVI 827
Db        477 CTIPAVCLLTGKFIIPTLSNLTSVWFLALFLSIIATGVLELRWSGVSIQDWWRNEQFWVI 536

Qy        828 GGVSAHLFAVFQGFLKMLAGIDTNFTVTAKAAEDAEFGELYMVKWTTLLIPPTTLLIINM 887
Db        537 GGVSAHLFAVFQGLLKVLAGVDTNFTVTAKAADDTEFGELYLFKWTTLLIPPTTLIILNM 596

Qy        888 SG-CAGFSDALNKGYEAWGPLFGKVFFAFWVILHLYPFLKGLMGRQNLTPTIVVLWSVLL 946
Db        597 VGVVAGVSDAINNGYGSWGPLFGKLFFAFWVILHLYPFLKGLMGRQNRTPTIVVLWSILL 656

Qy        947 ASVFSLVWVKINPFVNK 963
Db        657 ASIFSLVWVRIDPFLPK 673
```

```
RESULT 10
US-09-838-586-7
; Sequence 7, Application US/09838586
; Patent No. 6576818
; GENERAL INFORMATION:
;  APPLICANT: Stalker, D. et al.
;  TITLE OF INVENTION: Plant Cellulose Synthase and Promoter
;  TITLE OF INVENTION:  Sequences
;  FILE REFERENCE: 15621/02/US
;  CURRENT APPLICATION NUMBER: US/09/838,586
;  CURRENT FILING DATE:  2001-04-18
;  PRIOR APPLICATION NUMBER: 60/029,987
;  PRIOR FILING DATE: 1996-10-29
;  PRIOR APPLICATION NUMBER: 08/960,048
;  PRIOR FILING DATE: 1997-10-29
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7
;   LENGTH: 685
;   TYPE: PRT
;   ORGANISM: Gossypium hirsutum
US-09-838-586-7

  Query Match             52.2%;  Score 2733;  DB 4;  Length 685;
  Best Local Similarity   75.2%;  Pred. No. 3.8e-255;
  Matches  509;  Conservative   66;  Mismatches   74;  Indels   28;  Gaps    8;

Qy        311 ARKWVPFCKKFSIEPRAPEFYFSQKIDYLKDKVQPSFVKERRAMKRDYEEYKVRVNALVA 370
Db          1 ARRWVPFCKKHNVEPRAPEFYFNEKIDYLKDKVHPSFVKERRAMKREYEEFKVRINALVA 60

Qy        371 KAQKTPEEGWTMQDGTPWPGNNTRDHPGHDSGLPWEILGARDIEGNELPRLVYVSREKRP 430
Db         61 KAQKKPEEGWVMQDGTPWPGNNTRDHPGMIQ-VYLGSAGALDVDGKELPRLVYVSREKRP 119

Qy        431 GYQHHKKAGAENALVRVSAVLTNAPYILNVDCDHYVNNSKAVREAMCILMDPQVGRDVCY 490
Db        120 GYQHHKKAGAENALVRVSAVLTNAPFILNLDCDHYINNSKAMREAMCFLMDPQFGKKLCY 179

Qy        491 VQFPQRFDGIDKSDRYANRNVVFFDVNMKGLDGIQGPVYVGTGCVFNRQALYGYGPPSMP 550
Db        180 VQFPQRFDGIDRHDRYANRNVVFFDINMLGLDGLQGPVYVGTGCVFNRQALYGYDPP--V 237

Qy        551 SLRKRKDSSSCF----SCCCPSKKKPAQDPAE--------VYRDAKR-------EDLNAA 591
Db        238 SEKRPKMTCDCWPSWCCCCCCGGSRKKSKKKGEKKGLLGGLLYGKKKMMGKNYVKKGSAP 297

Qy        592 IFNLTEI----DNYDEHERSMLISQLSFEKTFGLSSVFIESTLMENGGVPESANSPPFIK 647
Db        298 VFDLEEIEEGLEGYEELEKSTLMSQKNFEKRFGQSPVFIASTLMENGGLPEGTNSTSLIK 357

Qy        648 EAIQVIGCGYEEKTEWGKQIGWIYGSVTEDILSGFKMHCRGWRSIYCMPVRPAFKGSAPI 707
Db        358 EAIHVISCGYEEKTEWGKEIGWIYGSVTEDILTGFKMHCRGWKSVYCVPKRPAFKGSAPI 417

Qy        708 NLSDRLHQVLRWALGSVEIFFSRHCPLWYGFGGGRLKWLQRLAYINTIVYPFTSLPLIAY 767
Db        418 NLSDRLHQVLRWALGSVEIFLSRHCPLWYGY-GGKLKWLERLAYINTIVYPFTSIPLLAY 476

Qy        768 CTIPAVCLLTGKFIIPTLSNLASMLFLGLFISIIVTAVLELRWSGVSIEDLWRNEQFWVI 827
Db        477 CTIPAVCLLTGKFIIPTLSNLTSVWFLALFLSIIATGVLELRWSGVSIQDWWRNEQFWVI 536

Qy        828 GGVSAHLFAVFQGFLKMLAGIDTNFTVTAKAAEDAEFGELYMVKWTTLLIPPTTLLIINM 887
Db        537 GGVSAHLFAVFQGLLKVLAGVDTNFTVTAKAADDTEFGELYLFKWTTLLIPPTTLIILNM 596

Qy        888 SG-CAGFSDALNKGYEAWGPLFGKVFFAFWVILHLYPFLKGLMGRQNLTPTIVVLWSVLL 946
Db        597 VGVVAGVSDAINNGYGSWGPLFGKLFFAFWVILHLYPFLKGLMGRQNRTPTIVVLWSILL 656

Qy        947 ASVFSLVWVKINPFVNK 963
```

```
        ||:||||||:|:||: |
      657 ASIFSLVWVRIDPFLPK 673


SULT 11
-09-221-013A-2
Sequence 2, Application US/09221013A
Patent No. 6495740
GENERAL INFORMATION:
 APPLICANT:  Arioli, Antonio
 APPLICANT:  Williamson, Richard E.
 APPLICANT:  Betzner, Andreas S.
 APPLICANT:  Peng, Liangcai
 TITLE OF INVENTION:  Manipulation of cellulose and/or Beta-1,4-glucan
 FILE REFERENCE:  96-98
 CURRENT APPLICATION NUMBER: US/09/221,013A
 CURRENT FILING DATE:  1998-12-23
 PRIOR APPLICATION NUMBER:  PCT/AU97/00402
 PRIOR FILING DATE:  1997-06-24
 PRIOR APPLICATION NUMBER:  AU PO0699
 PRIOR FILING DATE:  1996-06-27
 NUMBER OF SEQ ID NOS:  37
 SOFTWARE:  PatentIn Ver. 2.0
SEQ ID NO 2
  LENGTH: 629
  TYPE: PRT
  ORGANISM: Arabidopsis thaliana
-09-221-013A-2

Query Match           43.5%;  Score 2281.5;  DB 4;  Length 629;
Best Local Similarity  67.5%;  Pred. No. 1.6e-211;
Matches  426;  Conservative   73;  Mismatches  103;  Indels   29;  Gaps    9;

Qy      352 RAMKRDYEEYKVRVNALVAKAQKTPEEGWTMQDGTPWPGNNTRDHP-------GHDSGLP 404
            |||||:|||:|||:|||||||| | ||||||||||||||||||||       || ||
Db        1 RAMKREYEEFKVRINALVAKAQKIPGEGWTMQDGTPWPGNNTRDHPGMIQVFLGHSGGL- 59

Qy      405 WEILGARDIEGNELPRLVYVSREKRPGYQHHKKAGAENALVRVSAVLTNAPYILNVDCDH 464
                | :|||||||:|||||||||:|||||||| || :||||||||  |:|||||||
Db       60 -------DTDGNELPRLIYVSREKRPGFQHHKKAGAMNASIRVSAVLTNGAYLLNVDCDH 112

Qy      465 YVNNSKAVREAMCILMDPQVGRDVCYVQFPQRFDGIDKSDRYANRNVVFFDVNMKGLDGI 524
            | |||||::|||| :||| :|:  ||||||||||||||  |||||||:||||:||||||||
Db      113 YFNNSKAIKEAMCFMMDPAIGKKCCYVQFPQRFDGIDLHDRYANRNIVFFDINMKGLDGI 172

Qy      525 QGPVYVGTGCVFNRQALYGYGPPSMPSLRKRKDSSSCFSCCCPSKKKPAQDPAEVY---R 581
            |||||||||| |||||||| |       :  : :     || |:||          !   |
Db      173 HGPVYVGTGCCFNRQALYGYDP---VLTEEDLEPNIIVKSCCGSRKKGKSSKKYNYEKRR 229

Qy      582 DAKREDLNAAIFNLTEID----NYDEHERSMLISQLSFEKTFGLSSVFIESTLMENGGVP 637
              | | || :||: :||     ||: |||:|:|| | || || | ||| :| || ||:|
Db      230 GINRSDSNAPLFNMEDIDEGFEGYDD-ERSILMSQRSVEKRFGQSPVFIAATFMEQGGIP 288

Qy      638 ESANSPPFIKEAIQVIGCGYEEKTEWGKQIGWIYGSVTEDILSGFKMHCRGWRSIYCMPV 697
              : |    :|||| || ||||:||||||:|||||||||||||:||||| ||| |||| |
Db      289 PTTNPATLLKEAIHVISCGYEDKTEWGKEIGWIYGSVTEDILTGFKMHARGWISIYCNPP 348

Qy      698 RPAFKGSAPINLSDRLHQVLRWALGSVEIFFSRHCPLWYGFGGGRLKWLQRLAYINTIVY 757
            ||||||||||||||||:|||||||:||  |||||:|||:  |||: |:|:|||||||||
Db      349 RPAFKGSAPINLSDRLNQVLRWALGSIEILLSRHCPIWYGY-HGRLRLLERIAYINTIVY 407

Qy      758 PFTSLPLIAYCTIPAVCLLTGKFIIPTLSNLASMLFLGLFISIIVTAVLELRWSGVSIED 817
            | ||:|||||| :|| ||:| :|||| :|| ||: |: ||||| || :|:|:|:||||||
Db      408 PITSIPLIAYCILPAFCLITDRFIIPEISNYASIWFILLFISIAVTGILKLKWNGVSIED 467

Qy      818 LWRNEQFWVIGGVSAHLFAVFQGFLKMLAGIDTNFTVTAKAA-EDAEFGELYMVKWTTLL 876
             ||| |||||||  | ||||||||  ||:||||:||||||:|| :: :| :||: ||| ||
Db      468 WWRNNQFWVIGGTSTHLFAVFQGLLKVLAGINTNFTVTSKATNKNGDFAKLYIFKWTALL 527

Qy      877 IPPTTLLIINMSG-CAGFSDALNKGYEAWGPLFGKVFFAFWVILHLYPFLKGLMGRQNLT 935
            |||||:|::|: |  || | |:| ||::||||||:||| ||| ||||||||:|||| |
Db      528 IPPTTVLLVNLIGIVAGVSYAVNSGYQSWGPLFGKLFFALWVIAHLYPFLKGLLGRQNRT 587

Qy      936 PTIVVILWSVLLASVFSLVWVKINPFVNKVDN 966
                |||::||| |||:: ::||: ||| :  |
Db      588 PTIVIVWSVLLASIFSLLWVRINPFVDANPN 618


RESULT 12
US-09-221-013A-14
; Sequence 14, Application US/09221013A
; Patent No. 6495740
; GENERAL INFORMATION:
;  APPLICANT:  Arioli, Antonio
;  APPLICANT:  Williamson, Richard E.
;  APPLICANT:  Betzner, Andreas S.
;  APPLICANT:  Peng, Liangcai
;  TITLE OF INVENTION:  Manipulation of cellulose and/or Beta-1,4-glucan
;  FILE REFERENCE:  96-98
;  CURRENT APPLICATION NUMBER:  US/09/221,013A
;  CURRENT FILING DATE:  1998-12-23
;  PRIOR APPLICATION NUMBER:  PCT/AU97/00402
;  PRIOR FILING DATE:  1997-06-24
;  PRIOR APPLICATION NUMBER:  AU PO0699
;  PRIOR FILING DATE:  1996-06-27
;  NUMBER OF SEQ ID NOS:  37
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 14
;  LENGTH: 547
;  TYPE: PRT
;  ORGANISM: Oryza sativa
US-09-221-013A-14

  Query Match           26.7%;  Score 1396.5;  DB 4;  Length 547;
  Best Local Similarity  54.8%;  Pred. No. 5.3e-126;
  Matches  276;  Conservative   65;  Mismatches   94;  Indels   81;  Gaps   10;

Qy        8 ICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFFINEGRKVCLRCGSPV-------- 58
            :|  ||: ||  | ::|:||:| ::|:;: |:|:  ||   | :| : |
Db       40 VCQICGDTVGVSATGDVFVACNECAFPVCRPCYEYERKEGNQCCPQCKTRYKRHKGSPRV 99

Qy       59 ----DENLLGDVE---KKGSG-NQSTMASHLNDSQDVGIHARH-- -   ISSVSTVD 101
                :|  ::|::   | ':|      : |:   :||      ::|  :
Db      100 QGDEEEEDVDDLDNEFHYKHGNGKGPEWQIQRQGEDVDLSSSSRHEQHRIPRLTSGQQIS 159

Qy      102 SEMND --- ---      ---          - - -  -EYG NPI-WKNRVKSCKDKEN 148
                 |: |                            '| | : |: |; . .:|::
Db      160 GEIPDASPDRHSIA-DTSSYVDPSVPVPVPIVDPSKDLNSYGINSVDWQERVASWRNKQD 219

Qy      126 KK----KKRSPKAETRPAQVPTEQQMPFKPSAEASEPLSIVYPIPRNKLTPYRAVIIMRL 181
            |       : |:|    :          : :   ::  ||| : ||| ::!  || !!|:||
Db      220 KNMMQVANKYPPAPGDMEGTGSNGEDIQMVDQARLPLSRIVPIPSNQLNLYRIVIILRL 279

Qy      182 VILGLFFHFRITNPVDSAFGLWLTSVICEIWFAFSWVLDQFPKWNPVNRETYIERLSARY 241
            :||  || : :|:    |:|||| ||||| :|:|||:|:|||:|:|:|:|:|:|:|:|:|
Db      280 IILMFFPQYRVTHPVRDAYGLWLVSVICEIWLPLSWILDQFPKWYPINRETYLDRLALRY 339

Qy      242 EREGEPSQLAGVDFFVSTVDPLKEPPLITANTVLSILAVDYPVDKVSCYVSDDGAAMLSF 301
            :|||:|||  :|  |||| ||||:|||| |||||:|||||||||||||:|||:|:|:|
Db      340 DREGEPSQLAPIDVFVSTVDPLKEPPLITANTVLSILAVDYPVDKVSCYVSDDGSAMLTF 399

Qy      302 ESLVETAEFARKWVPFCKKFSIEPRAPEFYFSQKIDYLKDKVQPSFVKERRAMKREYEEY 361
            |:| |||||||||||||:| :| ||| :||:|:|||||:|:|||||||||||:|||::
Db      400 EALSETAEFARKWVPFCKKHNIEPRAPEFYFAQKIDYLKDKIQPSFVKERRAMKREYEEF 459

Qy      362 KVRVNALVAKAQKTPEEGWTMQDGTPWPGNNTRDHP-------GHDSGLPWEILGARDIE 414
            |||:||:|||: ||:|||  |||  |||:'||    ||  ||      : :
Db      460 KVRINALVAKAQKVPEEGWTMADGTAWPGNNPRDHPGMIQVFLGHSGGL - --- --DTD 511

Qy      415 GNELPRLVYVSREKPPGYQHHKKAGAENALVRVSAV 450
            |||||||||||||: ||:|||||'| ||':||||
Db      512 GNELPRLVYVSREKPPGFQHHKKAGAMNALIRVSAV 547


RESULT 13
US-08-960-048-11
```

```
; Sequence 11, Application US/08960348C
; Patent No. 6271443
; GENERAL INFORMATION:
;  APPLICANT:  Stalker, D. et al.
;  TITLE OF INVENTION:  Plant Cellulose Synthase and Promoter
;  TITLE OF INVENTION:  Sequences
;  FILE REFERENCE:  15621/01/US
;  CURRENT APPLICATION NUMBER: US/08/960,048C
;  CURRENT FILING DATE:  1997-10-29
;  PRIOR APPLICATION NUMBER:  60/029,987
;  PRIOR FILING DATE:  1996-10-29
;  NUMBER OF SEQ ID NOS:  12
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 11
;   LENGTH: 693
;   TYPE: PRT
;   ORGANISM: Escherichia coli
US-08-960-048-11

  Query Match             4.4%;  Score 228.5;  DB 3;  Length 693;
  Best Local Similarity   18.1%;  Pred. No. 8.3e-13;
  Matches  146;  Conservative   86;  Mismatches  229;  Indels  345;  Gaps   27;

Qy        176 VIIMRLVILGLFFHFRIT---   NPVDSAFGLWLTSVICEIWFAFSWVLDQFPKWNPVN 229
Db         22 LIVLSLTVSCRYIWWRYTSTLNWFDPVSLVCGLILLFAITYAWIVL--VLGYFQVVWPLN 79

Qy        230 RETYIERLSARYEREGEPSQLAGVDFFVSTVDPLKEPPLITANTVLSILAVDYPVDKVSC 289
Db         80 RQP---  ---VPLPKDMSLWPSVDIFVPTYN--EDLNVVKNTIYASLGIDWPKDKLNI 128

Qy        290 YVSDDGAAMLSFESLVETAEFARKWVPFCKKFSIEPRAPEFYFSQKIDYLKDKVQPSFVK 349
Db        129 WILDDGG--------  ----  -  --  ---  - --------------- 135

Qy        350 ERRAMKRDYEEYKVRVNALVAKAQKTPEEGWTMQDGTPWPGNNTRDHPGHDSGLPWEILG 409
Db        136 -----REEFRQFAQNVG- ---- -- --- --  ----- ---------------- 147

Qy        410 ARDIEGNELPRLVYVSREKRPGYQHHKKAGAENALVRVSAVLTNAPYILNVDCDHYVNNS 469
Db        148 ----- -----VKYIARTT-- HEHAKAGNINNAL--  -KYAKGEFVSIFDCDHVPTRS 188

Qy        470 KAVREAMCILMDPQVGRDVCYVQFPQRFDGIDKSDR    -  YANRNVVFFDVNMKGLD 522
Db        189 FLQMTMGWFLKEKQ ---LAMMQTPHHFFSPDPFERNLGRFRKTPNEGTLFYGLVQDGND 244

Qy        523 GIQGPVYVGTGCVFNRQALYGYGPPSMPSLRKRKDSSSCFSCCCPSKKKPAQDPAEVYRD 582
Db        245 MWDATFFCGS---  ----  -  --  - --  - AVIRRKP---------- 262

Qy        583 AKREDLNAAIFNLTEIDNYDEHERSMLISQLSFEKTFGLSSVFIESTLMENGGVPESANS 642
Db        263 ---  ----------------  -  --  - --  -  -LDEIGGI------ 269

Qy        643 PPFIKEAIQVIGCGYEEKTEWGKQIGWIYGSVTEDILSGFKMHCRGWRSIYCMPVRPAFK 702
Db        270  -----AVE--------- ---  -------TVTEDAHTSLRLHRRGYTSAY-MRI-PQAA 300

Qy        703 GSAPINLSDRLHQVLRWALGSVEIFFSRHCPLWYGFGGGRLKWLQRLAYINTIVYPFTSL 762
Db        301 GLATESLSAHIGQRIRWARGMVQIFRLDN-PL----TGKGLKFAQRLCYVNAMFHFLSGI 355

Qy        763 PLIAYCTIPAVCLLTGKFIIPTLSNLASMLFLGLFISIIVTAVLELRWSGVSIEDLWRNE 822
Db        356 PRLIFLTAPLAFLLLHAYII ----YAPALMIALF---VLFPMIHASLTNSKIQGKYRH- 406

Qy        823 QFWVIGGVSAHLFAVFQGFLKMLAGIDTNFTVTAKAADEAEFGELYMVKWTTLLIPPTTL 882
Db        407 SFW---  ---  --  ----  ----  --  -SEIYETVLAWYIAPPTLV 427

Qy        883 LIINMSGCAGFSDALNKGYEAWGPLFGKVFFAFWVILHLYPFLKGLMGRQNLTPTIVVLW 942
```

```
Db        428 ALINPH--KGKFNVTAKG----GGLVEEE-YVDWVISRPYIFLVLL----NLVGVAVGIW 476

Qy        943 -------SVLLASVFSLVWVKINPFV 961
Db        477 RYFYGPPTEMLTVVVSMVWVFYNLIV 502


RESULT 14
US-09-838-586-11
; Sequence 11, Application US/09838586
; Patent No. 6576818
; GENERAL INFORMATION:
;  APPLICANT: Stalker, D. et al.
;  TITLE OF INVENTION: Plant Cellulose Synthase and Promoter
;  TITLE OF INVENTION:  Sequences
;  FILE REFERENCE: 15621/02/US
;  CURRENT APPLICATION NUMBER: US/09/838,586
;  CURRENT FILING DATE:  2001-04-18
;  PRIOR APPLICATION NUMBER: 60/029,987
;  PRIOR FILING DATE: 1996-10-29
;  PRIOR APPLICATION NUMBER: 08/960,048
;  PRIOR FILING DATE: 1997-10-29
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11
;   LENGTH: 693
;   TYPE: PRT
;   ORGANISM: Escherichia coli
US-09-838-586-11

  Query Match             4.4%;  Score 228.5;  DB 4;  Length 693;
  Best Local Similarity   18.1%;  Pred. No. 8.3e-13;
  Matches  146;  Conservative   86;  Mismatches  229;  Indels  345;  Gaps   27;

Qy        176 VIIMRLVILGLFFHFRIT------NPVDSAFGLWLTSVICEIWFAFSWVLDQFPKWNPVN 229
Db         22 LIVLSLTVSCRYIWWRYTSTLNWDDPVSLVCGLILLFAITYAWIVL--VLGYFQVVWPLN 79

Qy        230 RETYIERLSARYEREGEPSQLAGVDFFVSTVDPLKEPPLITANTVLSILAVDYPVDKVSC 289
Db         80 RQP-------VPLPKDMSLWPSVDIFVPTYN---EDLNVVKNTIYASLGIDWPKDKLNI 128

Qy        290 YVSDDGAAMLSFESLVETAEFARKWVPFCKKFSIEPRAPEFYFSQKIDYLKDKVQPSFVK 349
Db        129 WILDDGG----------------------------------------------- --- 135

Qy        350 ERRAMKRDYEEYKVRVNALVAKAQKTPEEGWTMQDGTPWPGNNTRDHPGHDSGLPWEILG 409
Db        136 -----REEFRQFAQNVG-------------------------------------- 147

Qy        410 ARDIEGNELPRLVYVSREKRPGYQHHKKAGAENALVRVSAVLTNAPYILNVDCDHYVNNS 469
Db        148 -----------VKYIARTT---HEHAKAGNINNAL-----KYAKGEFVSIFDCDHVPTRS 188

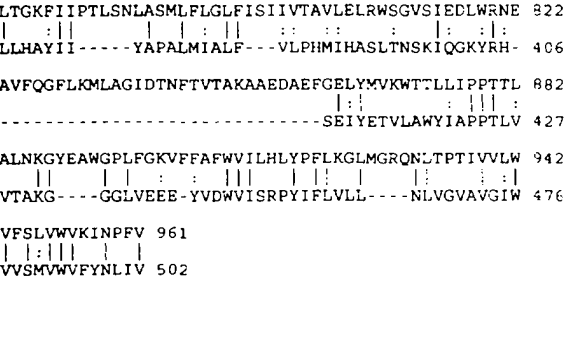
Qy        470 KAVREAMCILMDPQVGRDVCYVQFPQRFDGIDKSDR-------YANRNVVFFDVNMKGLD 522
Db        189 FLQMTMGWFLKEKQ----LAMMQTPHHFFSPDPFERNLGRFRKTPNEGTLFYGLVQDGND 244

Qy        523 GIQGPVYVGTGCVFNRQALYGYGPPSMPSLRKRKDSSSCFSCCCPSKKKPAQDPAEVYRD 582
Db        245 MWDATFFCGS------------------------------CAVIRRKP---------- 262

Qy        583 AKREDLNAAIFNLTEIDNYDEHERSMLISQLSFEKTFGLSSVFIESTLMENGGVPESANS 642
Db        263 ---------------------------------------------LDEIGGI------ 269

Qy        643 PPFIKEAIQVIGCGYEEKTEWGKQIGWIYGSVTEDILSGFKMHCRGWRSIYCMPVRPAFK 702
Db        270 ------AVE--------------------TVTEDAHTSLRLHRRGYTSAY-MRI-PQAA 300

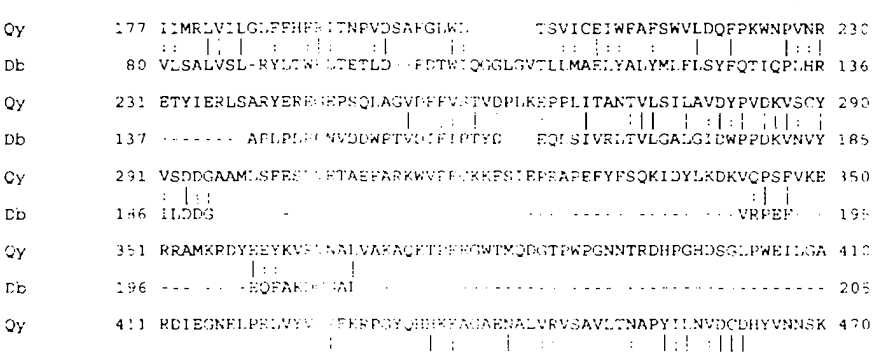
Qy        703 GSAPINLSDRLHQVLRWALGSVEIFFSRHCPLWYGFGGGRLKWLQRLAYINTIVYPFTSL 762
Db        301 GLATESLSAHIGQRIRWARGMVQIFRLDN-PL----TGKGLKFAQRLCYVNAMFHFLSGI 355
```



SULT 15
-07-689-008-2
Sequence 2, Application US/07689008
Patent No. 5268274
GENERAL INFORMATION:
APPLICANT: Ben-Bassat, Arie
APPLICANT: Calhoon, Roger D
APPLICANT: Fear, Anna L
APPLICANT: Gelfand, David H
APPLICANT: Meade, James H
APPLICANT: Tal, Rony
APPLICANT: Wong, Hing
APPLICANT: Benziman, Moshe
TITLE OF INVENTION: METHODS AND NUCLEIC ACID SEQUENCES FOR THE
TITLE OF INVENTION: EXPRESSION OF CELLULOSE SYNTHASE OPERON
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: McCutchen, Doyle, Brown & Enersen
STREET: Three Embarcadero Center
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/689,008
FILING DATE: 19910422
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 337,194
FILING DATE: 12-APR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 496,236
FILING DATE: 23-MAR-1990
ATTORNEY/AGENT INFORMATION:
NAME: Murphy, Lisabeth Feix
REGISTRATION NUMBER: 31547
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 393-2000
TELEFAX: (415) 393-2286
TELEX: 340817 MACPAG SFO
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 3031 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
S-07-689-008-2

Query Match 4.3%; Score 223; DB 1; Length 3031;
Best Local Similarity 19.9%; Pred. No. 3.6e-11;

Matches 139; Conservative 62; Mismatches 210; Indels 284; Gaps 25;



Search completed: October 29, 2003, 18:32:03
Job time : 34.3752 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 29, 2003, 18:24:13 ; Search time 76.6132 Seconds
(without alignments)
2026.213 Million cell updates/sec

Title: US-09-980-043-2
Perfect score: 5240
Sequence: 1 MMESGAPICHTCGEQVGHDA.....PFVNKVLNTLVAETCISIDC 978

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No	Score	% Query Match	Length	DB	ID	Description
1	4904	93.6	946	22	AAB37893	Populus tremuloide
2	4532	86.5	974	19	AAW73308	Cellulose synthase
3	4524	86.3	974	19	AAW60033	Cotton cellulose s
4	4285.5	81.8	955	21	AAB16338	Eucalyptus grandis
5	4202	80.2	958	23	ABB92981	Herbicidally activ
6	3492.5	66.7	1039	19	AAW73304	Cellulose synthase
7	3470	66.2	1043	23	ABB93703	Herbicidally activ
8	3367	64.3	1039	21	AAY58837	Soybean cellulose
9	3366.5	64.2	1079	23	AAU10496	Corn cellulose syn
10	3359.5	64.1	1077	21	AAY84108	Amino acid sequenc
11	3359.5	64.1	1077	21	AAY84114	Amino acid sequenc
12	3356.5	64.1	1076	21	AAY84121	DNA encoding a mai
13	3352	64.0	1148	21	AAY58832	Corn cellulose syn
14	3350	63.9	1065	21	AAG48733	Arabidopsis thalia
15	3350	63.9	1065	23	ABB93356	Herbicidally activ
16	3349	63.9	1054	21	AAG48734	Arabidopsis thalia
17	3347	63.9	1081	19	AAW33817	Arabidopsis cellul
18	3347	63.9	1081	23	ABB93199	Herbicidally activ
19	3346	63.9	1026	23	ABB93522	Herbicidally activ
20	3342	63.8	1058	21	AAY84107	Amino acid sequenc
21	3337	63.7	1081	19	AAW33820	Arabidopsis cellul
22	3335.5	63.7	1075	21	AAY84109	Amino acid sequenc
23	3335.5	63.7	1075	21	AAY84115	Amino acid sequenc
24	3335.5	63.7	1075	21	AAY84120	Amino acid sequenc
25	3332.5	63.6	1086	21	AAY84111	Amino acid sequenc
26	3332.5	63.6	1086	21	AAY84116	Amino acid sequenc
27	3332.5	63.6	1086	21	AAY58835	Corn cellulose syn
28	3332	63.6	1074	21	AAY84110	Amino acid sequenc
29	3332	63.6	1074	21	AAY84112	Amino acid sequenc
30	3332	63.6	1074	21	AAY84117	Amino acid sequenc
31	3332	63.6	1094	21	AAY84113	Amino acid sequenc
32	3332	63.6	1094	21	AAY84118	Amino acid sequenc
33	3332	63.6	1165	21	AAY58834	Corn cellulose syn
34	3329	63.5	1065	19	AAW33819	Arabidopsis cellul
35	3278	62.6	1069	23	ABB93411	Herbicidally activ
36	3276.5	62.5	1043	21	AAY84119	Amino acid sequenc
37	3274	62.5	1084	23	ABB93304	Herbicidally activ
38	3251	62.0	1065	23	ABB91880	Herbicidally activ
39	3248	62.0	1088	23	ABB91831	Herbicidally activ
40	3242	61.9	1084	19	AAW33818	Arabidopsis cellul
41	3238	61.8	959	21	AAG48735	Arabidopsis thalia
42	3226.5	61.6	1084	23	ABB93949	Herbicidally activ
43	3183	60.7	1080	22	AAB37895	Arabidopsis thalia
44	3156	60.2	821	23	AAU10495	Corn cellulose syn
45	3056	58.3	793	21	AAY58839	Soybean cellulose

ALIGNMENTS

RESULT 1
AAB37893
ID AAB37893 standard; Protein; 946 AA.
XX
AC AAB37893;
XX
DT 07-MAR-2001 (first entry)
XX
DE Populus tremuloides cellulose synthase CelA.
XX
KW Aspen; CelA; cellulose synthase; arabidopsis; wood;
KW mechanical stress response element; MSRE; cellulose biosynthesis.
XX
OS Populus tremuloides.
XX
PN WO200071670-A2.
XX
PD 30-NOV-2000.
XX
PF 18-MAY-2000; 2000WO-US13637.
XX
PR 21-MAY-1999; 99US-0135280.
XX
PA (UNMT) UNIV MICHIGAN TECHNOLOGICAL.
XX
PI Chiang VL, Wu L, Chandrashekhar JP, Carraway DT;
XX
DR WPI; 2001-032025/04.
DR N-PSDB; AAC65448.
XX
PT Novel polynucleotide encoding cellulose synthase useful for enhancing
PT cellulose biosynthesis, modifying lignin biosynthesis, and for

accelerating growth of plants and for upregulating cellulose synthase levels -

Claim 8; Fig 1; 65pp; English.

The present sequence is given in a specification relating to polynucleotides encoding cellulose synthase enzymes from Populus tremuloides and Arabidopsis. The cellulose synthases are designated PtCelA and AraxCelA respectively. The polynucleotides may be introduced into plants where they are useful for altering the growth of a plant. Vectors comprising a cellulose synthase promoter, or a fragment of the promoter containing one or more mechanical stress response elements (MSRE), operably linked to a cellulose synthase gene are useful for causing stress-induced gene expression in a plant cell. A cDNA encoding a protein that binds to a positive MSRE may be delivered into the plant to increase expression of cellulose synthase. A cDNA in an antisense orientation may be delivered to reduce expression of cellulose synthase. The polynucleotides enable the development of transgenic tree varieties having increased cellulose content, decreased lignin content and therefore improved wood fibre characteristics. The cellulose synthase promoters are useful for synthesising greater quantities of high-crystalline cellulose in plant and preferably in trees, which permits an increase in proportion of cellulose in transgenic plants, greater strength of juvenile wood and fiber and acceleration of overall growth rate.

Sequence 946 AA;

```
Query Match             93.6%;  Score 4904;  DB 22;  Length 946;
Best Local Similarity   96.7%;  Pred. No. 0;
Matches  946;  Conservative    0;  Mismatches    0;  Indels   32;  Gaps   16;

Qy          1 MMESGAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEINEGRKVCLRCGSPYDE 60
              ||||||||||||||  ||||||||||||||  ||||||.|||||||  |||||||.||||
Db          1 MMESGAPICHTCGE--GHDANGELFVACHE--YPMCKSCFEFEINE--KVCLRCGSPYDE 54

Qy         61 NLLDDVEKKGSGNQSTMASHLNDSQDVGIHARHISSVSTVDSEMNDEYGNPIWKNRVKSC 120
              ||  |||||||||||||  ||||||||||||||  |||||||||||||  ||||||||
Db         55 NL--DVEKKGSGNQSTMA--LNDSQDVGIHARHI--VSTVDSEMNDEYGN--WKNRVKSC 106

Qy        121 KDKENKKKKRSPKAETEPAQVPTEQQMEEKPSAEASEPLSIVYPIPRNKLTPYRAVIIMR 180
              |||||  ||||||||||||||  ||||||||||||||  ||||||||||||||  ||||
Db        107 KDKEN--KKRSPKAETEPAQVP--QQMEEKPSAEASEP--IVYPIPRNKLTPYR--IIMR 158

Qy        181 LVILGLFFHFRITNPVDSAFGLWLTSVICEIWFAFSWVLDQFPKWNPVNRETYIERLSAR 240
              ||||||||||  ||||||||||||||  ||||||||||||||  ||||||||||||||
Db        159 LVILGLFFHF--TNPVDSAFGLWLTS--CEIWFAFSWVLDQF--WNPVNRETYIERLS-- 210

Qy        241 YEREGEPSQLAGVDFFVSTVDPLKEPPLITANTVLSILAVDYPVDKVSCYVSDDGAAMLS 300
              ||||||||||||||  ||||||||||||||||||||||||||||||||||||||||||||
Db        211 YEREGEPSQLAGVD--VSTVDPLKEPPLITANTVLSILAVDYPVDKVSCYVSDDGAAMLS 268

Qy        301 FESLVETAEFARKWVPFCKKFSIEPRAPEFYFSQKIDYLKDKVQPSFVKERRAMKRDYEE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        269 FESLVETAEFARKWVPFCKKFSIEPRAPEFYFSQKIDYLKDKVQPSFVKERRAMKRDYEE 328

Qy        361 YKVRVNALVAKAQKTPEEGWTMQDGTPWPGNNTRDHPGHDSGLPWEILGARDIEGNELPR 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        329 YKVRVNALVAKAQKTPEEGWTMQDGTPWPGNNTRDHPGHDSGLPWEILGARDIEGNELPR 388

Qy        421 LVYVSREKRPGYQHHKKAGAENALVRVSAVLTNAPYILNVDCDHYVNNSKAVREAMCILM 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        389 LVYVSREKRPGYQHHKKAGAENALVRVSAVLTNAPYILNVDCDHYVNNSKAVREAMCILM 448

Qy        481 DPQVGRDVCYVQFPQRFDGIDKSDRYANRNVVFFDVNMKGLDGIQGPVYVGTGCVFNRQA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        449 DPQVGRDVCYVQFPQRFDGIDKSDRYANRNVVFFDVNMKGLDGIQGPVYVGTGCVFNRQA 508

Qy        541 LYGYGPPSMPSLRKRKDSSSCFSCCCPSKKKPAQDPAEVYRDAKREDLNAAIFNLTEIDN 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        509 LYGYGPPSMPSLRKRKDSSSCFSCCCPSKKKPAQDPAEVYRDAKREDLNAAIFNLTEIDN 568

Qy        601 YDEHERSMLISQLSFEKTFGLSSVFIESTLMENGGVPESANSPPFIKEAIQVIGCGYEEK 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        569 YDEHERSMLISQLSFEKTFGLSSVFIESTLMENGGVPESANSPPFIKEAIQVIGCGYEEK 628

Qy        661 TEWGKQIGWIYGSVT[illegible]LSGFFMPHPWRSIYGMPVRPAFKGSAPINLSDRLHQVLRWA 720
              ||||||||||||||[illegible]||||||||||||||||||||||||||||||||||||
Db        629 TEWGKQIGWIYGSVT[illegible]LSGFFMPHPWRSIYGMPVRPAFKGSAPINLSDRLHQVLRWA 688

Qy        721 LGSVEIFFSRHCPLWFGSRGRLKWLQRLAYINTIVYPFTSLPLIAYCTIPAVCLLTGKF 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        689 LGSVEIFFSRHCPLWFGSRGRLKWLQRLAYINTIVYPFTSLPLIAYCTIPAVCLLTGKF 748

Qy        781 IIPTLSNLASMLFLSIFISIIVTAVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQG 840
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        749 IIPTLSNLASMLFLSIFISIIVTAVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQG 808

Qy        841 FLKMLAGIDTNFTVTSKAADEAEFGELYMVKWTTLLIPPTTLLIINMSGVAGFSDAINKG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        809 FLKMLAGIDTNFTVTSKAADEAEFGELYMVKWTTLLIPPTTLLIINMSGVAGFSDAINKG 868

Qy        901 YEAWGPLFGKVFFAFWVILHLYPFL[illegible]LMGRQNRTPTIVVLWSVLLASVFSLVWVKINPF 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        869 YEAWGPLFGKVFFAFWVILHLYPFL[illegible]LMGRQNRTPTIVVLWSVLLASVFSLVWVKINPF 928

Qy        961 VNKVDNTLVAET[illegible]
              ||||||||||||
Db        929 VNKVDNTLVAET[illegible]
```

RESULT 2
AAW73308
ID AAW73308 standard; Prot; 974 AA.
XX
AC AAW73308;
XX
DT 08-FEB-1999 (first entry)
XX
DE Cellulose synthase PcwA1
XX
KW Cellulose synthase; cotton.
XX
OS Gossypium hirsutum.
XX
PN EP875575-A2.
XX
PD 04-NOV-1998.
XX
PF 31-MAR-1998; 98EP-0302489
XX
PR 01-APR-1997; 97JP-00081133.
XX
PA (NISN) NISSHINBO IND INC.
XX
PI Aotsuka S, Hasegawa O, Hayashi T, Ihara Y;
XX
DR WPI; 1998-559440/48.
DR N-PSDB; AAV08372.
XX
PT New DNA encoding cotton cellulose synthase proteins - useful for
PT controlling cellulose synthesis in prokaryotic or eukaryotic cells
XX
PS Claim 1; Page 14-18; 51pp; English
XX
CC This sequence is a cellulose synthase of the invention. The DNA can
CC be used for controlling cellulose synthesis in a cell by introducing the
CC DNA into the cell, expressing RNA with a nucleotide sequence homologous
CC to the DNA or a nucleotide sequence complementary to the DNA. The cells
CC are prokaryotic cells i.e. acetobacterium and/or eukaryotic cells such as
CC yeasts i.e. Saccharomyces spp., plant cells (cotton plants) and cultured
CC cells of mammals etc.
XX
SQ Sequence 974 AA;

Query Match 86.5%; Score 4532; DB 19; Length 974;

```
 Best Local Similarity   85.5%;  Pred. No. 0;
 Matches  840;  Conservative   72;  Mismatches   58;  Indels   12;  Gaps    7;

Qy          1 MMESGAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEINEGRKVCLRCGSPYDE 60
              |||||  |:|||||| || : ||| |:||||||::|:||||||::: ||:' ||||| ||:|
Db          1 MMESGVPVCHTCGEHVGLNVNGEPFVACHECNFPICKSCFEYDLKEGQKACLRCGIPYDE 60

Qy         61 NLLDDVEKKGSGNQSTMASHLNDSQDVGIHARHISSVSIVCSEMNDEYGNPIWKNRVKSC 120
              |||||:|| | :|:|:||||:|:: ||||'||||||||||:|||| :: :|||||||||:|
Db         61 NLLDDVE-KATGDQSTMAAHLSKSQDVGIHARHISSVSTICSEMTEDNGNPIWKNRVESW 119

Qy        121 KDKENKKKKRSPKAETEPAQVPTEQQMEEKPSAEASEPLSIVYPIPRNKLTPYRAVIIMR 180
              |:|:|||||  :      |::| ||||:|:|: :||:||:. .|||:::| ||| ||||| 
Db        120 KEKKNKKKKPATTKVEREAEIPPEQQMEDKPAPDASQPLSIIPIPKSRLAPYRTVIIMR 179

Qy        181 LVILGLFFHFRITNPVDSAFGLWLTSVICEIWFAFSWVLDQFPKWNPVNRETYIERLSAR 240
              |:||||||:|:||||||||'|||||' |||:|'||||||||||'||| :|||||||:|||| 
Db        180 LIILGLFFHYRVTNPVDSAFGLWLTSVICEIWFAFSWVLDQFPKWYPVNRETYIDRLSAR 239

Qy        241 YEREGEPSQLAGVDFFVSTVDPLKEPPLITANTVLSILAVDYPVDKVSCYVSDDGAAMLS 300
              |||||||::|| |||||||| ||||'|||||:|||||||:::|||'||:|||||||:   
Db        240 YEREGEPNELAAVDFFVSTVDPLKEPPLITANTVLSILALDYPVDKVSCYISDDGAAMLT 299

Qy        301 FESLVETAEFARKWVPFCKKFSIEPRAPEFYFSQKIDYLKDKVQPSFVKERRAMKRDYEE 360
              |||||||:|||||:|||| .|||:|||||:|||||||||'|||  |||||||||||||| 
Db        300 FESLVETADFARKWVPFCKKFSIEPRAPEFYFSQKIDYLKDKVQPSFVKERRAMKRDYEE 359

Qy        361 YKVRVNALVAKAQKTPEEGWTMQDGTPWPGNNTRDHPGMISQVFLGYSGAHDIEGNE 417
              |:|:|||||||||||||||||,''|||:' || :|||     |:     |   || ||||| 
Db        360 YKIRINALVAKAQKTPEEGWTMQDGTPWPGNNPRDHP     GMIQVFLGYSGAHDIEGNE 415

Qy        418 LPRLVYVSREKRPGYQHHKKAGAENALVRVSAVLTNAPYILNVDCDHYVNNSKAVREAMC 477
              |||||||||||||||||||||||||:'|||,'||||||:|.|:|||||'||||||||||| 
Db        416 LPRLVYVSREKRPGYQHHKKAGAENALVRVSAVLTNAPFILNLDCDHYVNNSKAVREAMC 475

Qy        478 ILMDPQVGRDVCYVQFPQRFDGIDKSDRYANRNVVFFDVNMKGLDGIQGPVYVGTGCVFN 537
              |'|||||||||||||||||,''|:|:|||' ||||:'||||||||||||||||||||||| 
Db        476 SLMDPQVGRDVCYVQFPQRFDGIDRSDRYANRNTVFFDVNMKGLDGIQGPVYVGTGCVFN 535

Qy        538 RQALYGYGPPSMPSLRKRKDSSSCFSCCCPSKKKPAQDPAEVYRDAKREDLNAAIFNLTE 597
              |:|:|||||||||  |  |||: |||| ||:| :||: :|||||||:|:|||| | 
Db        536 RQALYGYGPPSMPSFPK-SSSSSC-SCCCPGKKEP KDPSELYRDAKREELDAAIFNLRE 592

Qy        598 IDNYDEHERSMLISQLSFEKTFGLSAVFIESTLMENGGVPESANSPFFIKEAIQVIGCGY 657
              |: |||:|:|||||| || |||| |||:' ||| ||| || ||' | |:|||
Db        593 IDNYDEYERSMLISQTSFEKTFGLSAVFIESTLMENGGVPESANSPFFIKEAIHVIGCGY 652

Qy        658 EEKTEWGKQIGWIYGSVTEDILTGFKMHCRGWRSIYCMPKRAAFKGSAPINLSDRLHQVL 717
              |||| |||:|||||||||' |:|| ||||:|||. |:| :|||||||||||||||||| 
Db        653 EEKTAWGKEIGWIYGSVTEDILTGFKMHCRGWRSIYCMPLRAAFKGSAPINLSDRLHQVL 712

Qy        718 RWALGSVEIFFSRHCPLWYGFGGGRLKWLQRLAYINTIVYPFTSLPLIAYCTIPAVCLLT 777
              |||:|||||| |||||||:|||||.|:|||,'|||:'|||| ,||||||::||:|||| 
Db        713 RWALGSVEIFLSRHCPLWYGFGGGRLKWLQRLAYINTIVYPFTSLPLIAYCSLPAICLLT 772

Qy        778 GKFIIPTLSNLASMLFLGLFISIIVTAVLELRWSGVSIEDIWRNEQFWVIGGVSAHLFAV 837
              .|||||||||||:|||||: |||||: |||||  |||,.|||||||||||||||||||| 
Db        773 GKFIIPTLSNLASVLFLGLFLSIIVTAVLELRWSGVSIEDIWRNEQFWVIGGVSAHLFAV 832

Qy        838 FQGFLKMLAGIDTNFTVTAKAASDAEFGELYMVKWTTLLIPPTTLLIINMSG-CAGFSDA 896
              |||||||||||||||||||:|||:|:|||||||||||||||||||||||| |  ||||||
Db        833 FQGFLKMLAGIDTNFTVTAKAADDAEFGELYIVKWTTLLIPPTTLLIVNMVGVVAGFSDA 892

Qy        897 LNKGYEAWGPLFGKVFFAFWVILHLYPFLKGLMGRQNLTPTIVVLWSVLLASVFSLVWVK 956
              |||'||||||||||||:|||'|||||'|||':||| ||||||'|||||||||||||||: 
Db        893 LNKGYEAWGPLFGKVFFSFWVILHLYPFLKGLMGRQNRTPTIVVLWSVLLASVFSLVWVR 952

Qy        957 INPFVNKVDNTLVAETCISIDC 978
              |||||:  |:| |:::|||||
Db        953 INPFVSTADSTTVSQSCISIDC 974

RESULT 3
```

```
AAW60039
ID   AAW60039 standard; Protein; 974 AA.
XX
AC   AAW60039;
XX
DT   28-SEP-1998  (first entry)
XX
DE   Cotton cellulose synthase CelA1.
XX
KW   Cellulose synthase; CelA1; cotton; forest tree; plant fibre;
KW   wood; transgenic plant; ss.
XX
OS   Gossypium hirsutum Acala SJ-2.
XX
FH   Key             Location/Qualifiers
FT   Region          175..295
FT                   /label= H-1
FT   Region          280..295
FT                   /label= U-1
FT   Region          296..420
FT                   /label= P-CR
FT   Region          421..539
FT                   /label= H-2
FT   Region          450..467
FT                   /label= U-2
FT   Region          540..658
FT                   /label= HVR
FT   Region          659..880
FT                   /label= H-3
FT   Region          667..688
FT                   /label= U-3
FT   Region          706..722
FT                   /label= U-4
FT   Modified-site   81
FT                   /note= "N-glycosylated"
FT   Modified-site   466
FT                   /note= "N-glycosylated"
FT   Modified-site   703
FT                   /note= "N-glycosylated"
FT   Modified-site   846
FT                   /note= "N-glycosylated"
FT   Misc difference 293
FT                   /note= "critical Asp residue"
FT   Misc difference 461
FT                   /note= "critical Asp residue"
FT   Misc difference 672
FT                   /note= "critical Asp residue"
FT   Peptide         710..714
FT                   /note= "QXXRW motif"
XX
PN   WO9818949-A2.
XX
PD   07-MAY-1998.
XX
PF   29-OCT-1997;   97WO-US19529.
XX
PR   29-OCT-1996;   96US-0029987
XX
PA   (CALJ ) CALGENE INC.
XX
PI   Delmer DP,  Pear JR,  Stalker DM;
XX
DR   WPI; 1998-272238/24.
DR   N-PSDB; AAV34432.
XX
PT   New isolated plant cellulose synthesis enzyme DNA - is used to
PT   develop products for producing modified plant fibre products,
PT   particularly for modifying cotton fibre and wood quality
XX
PS   Example 5; Fig 3A-C; 49pp; English.
XX
CC   This full-length polypeptide comprises cotton cellulose synthase
CC   CelA1.  Its amino acid sequence was deduced from a cDNA clone (see
```

C AAV60039) isolated from a cotton fibre cDNA library. Expression of
C CelA1 and CelA2 (see AAW60040) in developing fibre is approximately
C 500 times higher than the level of expression in other cotton
C tissues; approximately 1-2% of the 24 dpa mRNA is CelA mRNA.
C Sequence analysis of CelA1 and CelA2 showed regions of high homology
C (H-1, H-2 and H-3) between plant and bacteral proteins, as well as
C a region (P-CR) that is conserved among plant CelA genes, and
C another region (HVR) that is hypervariable between plant genes.
C Regions U-1 to U-4 are predicted to be critical for UDP-glc binding
C and catalysis in bacterial CelA proteins. The isolated CelA cDNA
C clones are used in claimed methods for modifying fibre phenotype
C in a cotton plant and for modifying wood quality phenotype in a
C forest tree species. Lignin can be reduced by over-expressing the
C cellulose synthase product and diverting production into cellulose.
X
Q Sequence 974 AA;

Query Match 86.3%; Score 4524; DB 19; Length 974;
Best Local Similarity 85.5%; Pred. No. 0;
Matches 840; Conservative 70; Mismatches 60; Indels 12; Gaps 7;

```
Qy          1 MMESGAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEINEGRKVCLRCGSPYDE 60
              ||||| |:|||||| || : ||| |||||||::|:||||||::: |||| ||||.|||||
Db          1 MMESGVPVCHTCGEHVGLNVNGEPFVACHECNFPICKSCFEYDLKEGRKACLRCGSPYDE 60

Qy         61 NLLDDVEKKGSGNQSTMASHLNDSQDVGIHARHISSVSTVDSEMNDEYGNPIWKNRVKSC 120
              |||||| | :|:|||||:||| ||||||||||||||||:|||| :: || ||||||:|
Db         61 NLLDDVE-KATGDQSTMAAHLNKSQDVGIHARHISSVSTLDSEMAEDNGNSIWKNRVESW 119

Qy        121 KDKENKKKKRSPKAETEPAQVPTEQQMEEKPSAEASEPLSIVYPIPRNKLTPYRAVIIMR 180
              |:|:||||| :       |::| ||||||:|| :||:||| : |||:::| ||| |||||
Db        120 KEKKNKKKKPATTKVEREAEIPPEQQMEDKPAPDASQPLSTIIPIPKSRLAPYRTVIIMR 179

Qy        181 LVILGLFFHFRITNPVDSAFGLWLTSVICEIWFAFSWVLDQFPKWNPVNRETYIERLSAR 240
              |:|||||||:|:|||||||||||||||||||||||||||||:|||||| |||||||:|||||
Db        180 LIILGLFFHYRVTNPVDSAFGLWLTSVICEIWFAFSWVLDQFPKWYPVNRETYIDRLSAR 239

Qy        241 YEREGEPSQLAGVDFFVSTVDPLKEPPLITANTVLSILAVDYPVDKVSCYVSDDGAAMLS 300
              ||||||| :|| ||||||||||||||||||||||||||:||||||||||:||||||||:
Db        240 YEREGEPDELAAVDFFVSTVDPLKEPPLITANTVLSILALDYPVDKVSCYISDDGAAMLT 299

Qy        301 FESLVETAEFARKWVPFCKKFSIEPRAPEFYFSQKIDYLKDKVQPSFVKERRAMKRDYEE 360
              ||||||||:|||||||||||||||||||||||||||||||||||||||||||||||||||
Db        300 FESLVETADFARKWVPFCKKFSIEPRAPEFYFSQKIDYLKDKVQPSFVKERRAMKRDYEE 359

Qy        361 YKVRVNALVAKAQKTPEEGWTMQDGTPWPGNNTRDHPGHDSGLPWEIL---GARDIEGNE 417
              ||:|:|||||||||||:||||||||| ||||| ||||    |:    |   |||||||||
Db        360 YKIRINALVAKAQKTPDEGWTMQDGTSWPGNNPRDHP----GMIQVFLGYSGARDIEGNE 415

Qy        418 LPRLVYVSREKRPGYQHHKKAGAENALVRVSAVLTNAPYILNVDCDHYVNNSKAVREAMC 477
              ||||||||||||||||||||||||||||||||||||||:|||:|||||||||||||||||
Db        416 LPRLVYVSREKRPGYQHHKKAGAENALVRVSAVLTNAPFILNLDCDHYVNNSKAVREAMC 475

Qy        478 ILMDPQVGRDVCYVQFPQRFDGIDKSDRYANRNVVFFDVNMKGLDGIQGPVYVGTGCVFN 537
              |||||||||||||||||||||||:|||||||| ||||||||||||||||||||||||||
Db        476 FLMDPQVGRDVCYVQFPQRFDGIDRSDRYANRNTVFFDVNMKGLDGIQGPVYVGTGCVFN 535

Qy        538 RQALYGYGPPSMPSLRKRKDSSSCFSCCCPSKKKPAQDPAEVYRDAKREDLNAAIFNLTE 597
              |||||||||||||  |   |||| ||||| ||:| :||:|:|||||||:|:|||||| |
Db        536 RQALYGYGPPSMPSFPK-SSSSSC-SCCCPGKKEP-KDPSELYRDAKREELDAAIFNLRE 592

Qy        598 IDNYDEHERSMLISQLSFEKTFGLSSVFIESTLMENGGVPESANSPPFIKEAIQVIGCGY 657
              ||||||:|||||||| ||||||||||||||||||||| ||||    ||||| || |||
Db        593 IDNYDEYERSMLISQTSFEKTFGLSSVFIESTLMENGGVAESANPSTLIKEAIHVISCGY 652

Qy        658 EEKTEWGKQIGWIYGSVTEDILSGFKMHCRGWRSIYCMPVRPAFKGSAPINLSDRLHQVL 717
              |||| |||:|||||||||||||:||||||||||||||:|||||||||||||||||||||
Db        653 EEKTAWGKEIGWIYGSVTEDILTGFKMHCRGWRSIYCMPLRPAFKGSAPINLSDRLHQVL 712

Qy        718 RWALGSVEIFFSRHCPLWYGFGGGRLKWLQRLAYINTIVYPFTSLPLIAYCTIPAVCLLT 777
              |||||||||| |||||||||||||||||||||||||||||||||||||::||:||||
Db        713 RWALGSVEIFLSRHCPLWYGFGGGRLKWLQRLAYINTIVYPFTSLPLIAYCSLPAICLLT 772
```

```
Qy        778 GKFIIPTLSNLASVWFLGLFLSIIVTAVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAV 837
              |||:|||||||||:  |||:: |:|  |||| .|||.|||||||:|||||| ||||
Db        773 GKFIIPTLSNLASVIFLGLFLSIIVTAVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAV 832

Qy        838 FQGFLKMLAGIDTNFTVTAKAADDADFGELYMVKWTTLLIPPTTLLIINMSGVAGFSDA 896
              |||||||||||| ||:  |||-:|||:  |||:' ||||'|'||||:|:|| |  ||||!
Db        833 FQGFLKMLAGIDTNFTVTAKAADDADFGELYIVKWTTLLIPPTTLLIVNMVGVVAGFSDA 892

Qy        897 LNKGYEAWGPLFGKVFFAFWVILHLYPFLKGLMGRQNRTPTIVVLWSVLLASVFSLVWVK 956
              .||||:|||||:|:| :|||:'|||  '|||:'|| |:|:|||:'|||| '|||  :
Db        893 LNKGYEAWGPLFGKVFFSFWVILHLYPFLEGLMGRQNRTPTIVVLWSVLLASVFSLVWVR 952

Qy        957 INPFVNKVDNTLVAE 971
              |||||:  |:|  . :  .|||
Db        953 INPFVSTADSTFVN 967
```

RESULT 4
AAB16338
ID AAB16338 standard; P [illegible] 955 AA.
XX
AC AAB16338;
XX
DT 31-OCT-2000 (first [illegible]
XX
DE Eucalyptus grandis cellu[illegible] syn[illegible] protein sequence SEQ ID NO:146.
XX
KW Eucalyptus grandis; pinus radiata; Monterey pine; modification;
KW plant cell wall; polysaccharide; polysaccharide biosynthetic pathway;
KW transgenic plant.
XX
OS Eucalyptus grandis.
XX
PN WO200022092 A2.
XX
PD 20-APR-2000.
XX
PF 08-OCT-1999; 99WO-NZ00169.
XX
PR 13-OCT-1998; 98US-017862.
PR 11-AUG-1999; 99US-014426.
XX
PA (GENE-) GENESIS RES & DEV CORP LTD.
PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI Bloksberg LN;
XX
DR WPI; 2000-339328/29.
DR N-PSDB; AAA67145.
XX
PT New genes encoding proteins involved in a plant polysaccharide
PT biosynthetic pathway, useful for modulating or altering the
PT polysaccharide content, composition or structure of the plant -
XX
PS Claim 17; Page 111-115; 403pp; English.
XX
CC The present invention describes isolated polynucleotides (PN) comprising
CC a sequence selected from one of 835 nucleotide sequences given in
CC AAA67073 to AAA67907, their (reverse) complements, sequences producing
CC an Expectation (E) value of 0.01 or less compared to the 835 sequences,
CC sequences at least 50% identical to them, 200, 100, 40 or 20-mers of the
CC 835 sequences or sequences that are degenerately equivalent or allelic
CC to the 835 sequences. The polynucleotides are used to modify the
CC activity of a polypeptide involved in a polysaccharide biosynthetic
CC pathway in the plant. They are especially used to modulate or alter the
CC polysaccharide content, composition or structure of the plant. AAB16266
CC to AAB16340 are proteins encoded by some of the polynucleotide sequence
CC given in the present invention.
XX
SQ Sequence 955 AA;

Query Match 81.8%; Score 4285.5; DB 21; Length 955;
Best Local Similarity 83.2%; Pred. No. 0;

```
  Matches  799;  Conservative   73;  Mismatches   73;  Indels   15;  Gaps    7;

Qy          1 MMESGAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEINEGRKVCLRCGSPYDE 60
              |||||  |:|:|||| || |  ||:|||| ||:: ::|:|  :|| |||| ||||||:::
Db          1 MMESGVPLCNTCGEAVGVDEKGEVFVACQECNFAICKACVFYEIKEGRKACLRCGTPFEA 60

Qy         61 NLLDDVEKKGSGNQSTMASHLNDSQDVGIHARHISSVSTVDSEMNDEYGNPIWKNRVKSC 120
              | : | |:    |::|||: ||| || |||||||||||:|, ||| ||||||||:|
Db         61 NSMADAERNELGSRSTMAAQLNDPQDTGIHARHISSVSTIDSEYNDETGNPIWKNRVESW 120

Qy        121 KDKENKKKKRSPKAETEPAQVPTEQQMEEKPSAEASEPLSIVYPIPRNKLTPYRAVIIMR 180
              ||:|||,|   :|: | | :| |||:|,|| ,:||:|| . || ::|| ||| ||||
Db        121 KDKKNKKKKAPTKAEKE-AQVPPEQQMEEKQIADASEPLSIVIPIAKSKLAPYRTVIIMR 179

Qy        181 LVILGLFFHFRITNPVDSAFGLWLTSVICEIWFAFSWVLDQFPKWNPVNRETYIERLSAR 240
              |:|| ,|:|:|:|:|:||||: |||||:|'||||':||'''||,|||:|||| |:::|||||
Db        180 LIILALFFHYRVTHPVDSAYPLWLTSIICEIWFAYSWVLDQFPKWSPVNRITHVDRLSAR 239

Qy        241 YEREGEPSQLAGVDFFVSTVDPLKEPPLITANTVLSILAVDYPVDKVSCYVSDDGAAMLS 300
              ||:|||||:|| ||||||||||:|||||:|||||||||||||||||||:|||||||||
Db        240 YEKEGEPSELAAVDFFVSTVDPMKEPPLITANTVLSILAVDYPVDKVSCYLSDDGAAMLS 299

Qy        301 FESLVETAEFARKWVPFCKKFSIEPRAPEFYFSQKIDYLKDKVQPSFVKERRAMKRDYEE 360
              ||||||.:|||||||||||:||||||:|'|||||||||' |:|||||||||||||||||
Db        300 FESLVETADFARKWVPFCKKYSIEPRAPEFYFSQKIDYLKDKIQPSFVKERRAMKRDYEE 359

Qy        361 YKVRVNALVAKAQKTPEEGWTMQDGTPWPGNNTRDHPGHDSGLPWEIL---GARDIEGNE 417
              :|||||||||||| |||||:|:|||||  |:|||    '|    |  :| |||||||
Db        360 FKVRVNALVAKAQKAPEEGWSMQDGTPWPGNNSRDHP-   GMIQVFLGSSGAHDIEGNE 415

Qy        418 LPRLVYVSREKRPGYQHHKKAGAENALVRVSAVLTNAPYILNVDCDHYVNNSKAVREAMC 477
              |:|||||||||||:|||:||||'| |||.:|:|||||||:|:|||||||| | ||||||
Db        416 LPRLVYVSREKRPGFQHHKKAGAENALVRVSAILTNAPYILNLDCDHYVNYSNAVREAMC 475

Qy        478 ILMDPQVGRDVCYVQFPQRFDGIDKSDRYANRNVVFFDVNMKGLDGIQGPVYVGTGCVFN 537
              ,|:|||||::|||||||'|||:|'||| || ||| |||:|'||||'||||||||'|
Db        476 FLMDPQVGRNLCYVQFPQRFDGIDRSDRYANRNTVFFDVNMKGLDGIQGPVYVGTGCVFN 535

Qy        538 RQALYGYGPPSMPSLRKRKDSSS---CFGCCCPSKKKPAQDPAEVYRDAKREDLNAAIFN 594
              |||||||||||||:| |      '  ' |||:| ||: :| :|||||:||||||||||
Db        536 RQALYGYGPPSMPNLPKPSSSCSWCGCCSCCCPS-KKPTKDLSEVYRDSKREDLNAAIFN 594

Qy        595 LTEIDNYDEHERSMLISQLSFEKTFGLSSVFIESTLMENGGVPESANSPPFIKEAIQVIG 654
              | |||.|||||||||||:|||| |||  ||||  : |  |||::     |||| ||
Db        595 LGEIDNYDEHERSMLISQMSFEKTFGLSVVFIESTLIANGGVPESAHPSMLIKEAIHVIS 654

Qy        655 CGYEEKTEWCKQIGWIYGSVTEDILTGFKMHCRGWRSVYCMPKRPAFKGSAPINLSDRLH 714
              ||.|||| ||:|||||:|||| |||  |||:.|| ..|||_|||:|||||:||
Db        655 CGYEEKTAWGKEIGWIYGSVTEDIVTGFEMHCRGWRSVYCMPKRPAFKGSAPINLSDRLH 714

Qy        715 QVLRWALGSVEIFFSRHCPLWYGFGGGRLKWLQRLAYINTIVYPFTSLPLIAYCTIPAVC 774
              |||||||||||| |||| .|||| |||| |||:|||||||||||||||||||||:|||||:|
Db        715 QVLRWALGSVEIFLSRHCPLWYGFGGGRLKWLQRLAYINTIVYPFTSLPLVAYCTIPAIC 774

Qy        775 LLTGKFIIPTLSNLASMLFLGLFISIIVTAVLELRWSGVSIEDLWRNEQFWVIGGVSAHL 834
              |||||||||||||||:|||:|:|||||:||'||||'|||  ||||:|||||||||||:|
Db        775 LLTGKFIIPTLSNLASVLFLGLFLSIIVTSVLELRWSGVSIEDWWRNEQFWVIGGVSAHL 834

Qy        835 FAVFQGFLKMLAGIDTNFTVTAKAAEDAEFGELYMVKWTTLLIPPTTLLIINMSG-CAGF 893
              |:|||||||||:|||||| |:|:|| |||:|||:|| ||||||||:|| |  |||
Db        835 FAVFQGFLKMLAGLDTNFTVTTKAADDAEFGELYMIKWTTLLIPPTTLLIVNMVGVVAGF 894

Qy        894 SDALNKGYEAWGPLFGKVFFAFWVILHLYPFLKGLMGRQNRTPTIVVLWSV--LLASVFS 951
              |||||||||||||||||||:|||||:|||||'|||:|||  |||||||| ||:|: |
Db        895 SDALNKGYEAWGPLFGKVFFAFWVILHLYPFLKGLMGRQNRTPTIVVLWSVFWLLSSLSS 954


RESULT 5
ABB92981
ID   ABB92981 standard; Protein; 958 AA.
XX
AC   ABB92981;
XX
DT   31-MAY-2002  (first entry)
XX
DE   Herbicidally active polypeptide SEQ ID NO 2192.
XX
KW   Herbicidal; plant; agriculture; herbicide.
XX
OS   Arabidopsis thaliana.
XX
PN   WO200210210-A2.
XX
PD   07-FEB-2002.
XX
PF   28-AUG-2001; 2001WO-EP09892.
XX
PR   28-AUG-2001; 2001WO-EP09892.
XX
PA   (FARB ) BAYER AG.
XX
PI   Tietjen K,  Weidler M;
XX
DR   WPI; 2002-269010/31.
XX
PT   Identifying plant target proteins for herbicidally active compounds,
PT   comprising aligning and comparing nucleic acid or amino acid sequences
PT   from plant with nucleic acid or amino acid sequences from non-plant
PT   organisms -
XX
PS   Claim 5; SEQ ID NO 2192; 261pp + Sequence Listing; English.
XX
CC   The invention relates to identifying target proteins
CC   (ABB90790-ABB94016) for herbicidally active compounds, comprising
CC   aligning and comparing nucleic acid or amino acid sequences from plant
CC   with nucleic acid or amino acid sequences from non-plant organisms using
CC   suitable search parameters, where plant sequences having an E-value
CC   greater by a factor of 3 than the E-value of most similar non-plant
CC   sequences are selected. The polypeptides or nucleic acids encoding them
CC   are useful for identifying modulators. The identified modulators are
CC   useful as herbicides.
XX
SQ   Sequence   958 AA;

  Query Match             80.2%;  Score 4202;  DB 23;  Length 958;
  Best Local Similarity   79.1%;  Pred. No. 0;
  Matches  784;  Conservative   85;  Mismatches   76;  Indels   46;  Gaps    9;

Qy          1 MMESGAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEINEGRKVCLRCGSPYDE 60
              ,||| :|||:||||::|  :||| ||||||||:|:||:| |:|  |||::|||||:||||
Db          1 MMESRSPICNTCGEEIGVKSNGEFFVACHECSFPICKACLEYEFKEGRRICLRCGNPYDE 60

Qy         61 NLLDDVEKKGSGNQSTMASHLNDSQDVGIHARHISSVSTVDSEMNDEYGNPIWKNRVKSC 120
               |: |||| | |  ||          | |||||||||:|||:|||:|||||||||||:|
Db         61 NVFDDVETKTSKTQSI---------DSGIHARHISTVSTIDSELNDEYGNPIWKNRVESW 111

Qy        121 ---KDKENKKKKRSP---KAETEPAQVPTEQQMEEKP---SAEASEPLSIVYPIPRNKLT 171
                 |||::||||: |   |||   ||:||:| ||: |    : |:: ||:| |||| |:|
Db        112 KDKKDKKSKKKKKDPKATKAEQHEAQIPTQQHMEDTPPNTESGATDVLSVVIPIPRTKIT 171

Qy        172 PYRAVIIMRLVILGLFFHFRITNPVDSAFGLWLTSVICEIWFAFSWVLDQFPKWNPVNRE 231
                 || ||||||:|| |||::|||:|||||:|||||||||||||| ||||||||||:|:|||
Db        172 SYRIVIIMRLIILALFFNYRITHPVDSAYGLWLTSVICEIWFAVSWVLDQFPKWSPINRE 231

Qy        232 TYIERLSARYEREGEPSQLAGVDFFVSTVDPLKEPPLITANTVLSILAVDYPVDKVSCYV 291
              |||:|||||:||||| |||| ||||||||||||||||||||||||||:||||||||||
Db        232 TYIDRLSARFEREGEQSQLAAVDFFVSTVDPLKEPPLITANTVLSILALDYPVDKVSCYV 291

Qy        292 SDDGAAMLSFESLVETAEFARKWVPFCKKFSIEPRAPEFYFSQKIDYLKDKVQPSFVKER 351
              ||||:||||||||||:||||||||||:||||||||||| |||||:||||||||||
Db        292 SDDGAAMLSFESLVETADFARKWVPFCKKYSIEPRAPEFYFSLKIDYLRDKVQPSFVKER 351

Qy        352 RAMKRDYEEYKVRVNALVAKAQKTPEEGWTMQDGTPWPGNNTRDHPGHDSGLPWEIL--- 408
              |||||||||:|:|:|||||||||||||||||||| ||||||||||    |:    |
Db        352 RAMKRDYEEFKIRMNALVAKAQKTPEEGWTMQDGTSWPGNNTRDHP----GMIQVFLGYS 407
```

```
409 GARDIEGNELPRLVYVSREKRPGYQHHKKAGAENALVRVSAVLTNAPYILNVDCDHYVNN 468
408 GARDIEGNELPRLVYVSREKRPGYQHHKKAGAENALVRVSAVLTNAPFILNLDCDHYVNN 467

469 SKAVREAMCILMDPQVGRDVCYVQFPQRFDGIDKSDRYANRNVVFFDVNMKGLDGIQGPV 528
468 SKAVREAMCFLMDPVVGQDVCFVQFPQRFDGIDKSDRYANRNIVFFDVNMRGLDGIQGPV 527

529 YVGTGCVFNRQALYGYGPPSMPSLRKRKDSSSCFSCCCPSKKKPAQDPAEVYRDAKREDL 588
528 YVGTGTVFRRQALYGYSPPSKPRILPQSSSS---SCCCLTKKKQPQDPSEIYKDAKREEL 584

589 NAAIFNLTEIDNYDEHERSMLISQLSFEKTFGLSSVFIESTLMENGGVPESANSPPFIKE 648
585 DAAIFNLGDLDNYDEYDRSMLISQTSFEKTFGLSTVFIESTLMENGGVPDSVNPSTLIKE 644

649 AIQVIGCGYEEKTEWGKQIGWIYGSVTEDILSGFKMHCRGWRSIYCMPVRPAFKGSAPIN 708
645 AIHVISCGYEEKTEWGKEIGWIYGSITEDILTGFKMHCRGWRSIYCMPLRPAFKGSAPIN 704

709 LSDRLHQVLRWALGSVEIFFSRHCPLWYGFGGGRLKWLQRLAYINTIVYPFTSLPLIAYC 768
705 LSDRLHQVLRWALGSVEIFLSRHCPLWYGCSGGRLKLLQRLAYINTIVYPFTSLPLVAYC 764

769 TIPAVCLLTGKFIIPTLSNLASMLFLGLFISIIVTAVLELRWSGVSIEDLWRNEQFWVIG 828
765 TLPAICLLTGKFIIPTLSNLASMLFL----------GVSIEDLWRNEQFWVIG 807

829 GVSAHLFAVFQGFLKMLAGIDTNFTVTAKAAEDAEFGELYMVKWTTLLIPPTTLLIINMS 888
808 GVSAHLFAVFQGFLKMLAGLDTNFTVTSKTADDLEFGELYIVKWTTLLIPPTSLLIINLV 867

889 G-CAGFSDALNKGYEAWGPLFGKVFFAFWVILHLYPFLKGLMGRQNLTPTIVVLWSVLLA 947
868 GVVAGFSDALNKGYEAWGPLFGKVFFAFWVILHLYPFLKGLMGRQNRTPTIVILWSILLA 927

948 SVFSLVWVKINPFVNKVDNTLVAETCISIDC 978
928 SVFSLVWVRINPFVSKTDTTSLSLNCLLIDC 958
```

```
ESULT 6
AW73309
D   AAW73309 standard; Protein; 1039 AA.
X
C   AAW73309;
X
T   08-FEB-1999  (first entry)
X
E   Cellulose synthase PcsA2.
X
W   Cellulose synthase; cotton.
X
S   Gossypium hirsutum.
X
N   EP875575-A2.
X
D   04-NOV-1998.
X
F   31-MAR-1998;   98EP-0302489.
X
R   01-APR-1997;   97JP-0083133.
X
A   (NISN ) NISSHINBO IND INC.
X
I   Aotsuka S,  Hasegawa O,  Hayashi T,  Ihara Y;
X
R   WPI; 1998-559440/48.
R   N-PSDB; AAV08373.
X
T   New DNA encoding cotton cellulose synthase proteins - useful for
T   controlling cellulose synthesis in prokaryotic or eukaryotic cells
X
PS  Claim 1; Page 22-27; 51pp; English.
XX
CC  This sequence is a cellulose synthase of the invention. The DNA can
CC  be used for controlling cellulose synthesis in a cell by introducing the
CC  DNA into the cell, expressing RNA with a nucleotide sequence homologous
CC  to the DNA or a nucleotide sequence complementary to the DNA. The cells
CC  are prokaryotic cells i.e. acetobacterium and/or eukaryotic cells such as
CC  yeasts i.e. Saccharomyces spp., plant cells (cotton plants) and cultured
CC  cells of mammals etc.
XX
SQ  Sequence  1039 AA;

  Query Match            66.7%;  Score 4492.5;  DB 19;  Length 1039;
  Best Local Similarity  64.4%;  Pred. No. 0;
  Matches  663;  Conservative  122;  Mismatches  183;  Indels  61;  Gaps  14;

Qy          3 ESGAPICHTCGEQVATDRNGELFVACVECSYPMCKSCFEFEINEGRKVCLRCGSPY- -- 58
              :  :  :|:::  ||: :|||: ::|::: |:|:| :|| :  :| : :
Db         25 QSSTKICRVCGDRIQKENGQPFVACVECAFPVCRPCYEYERSEGNQCCPQCNTRYKRHK 84

Qy         59 ----- -  --  IPVLLEDVERPGSKPNQSTMASHLNDSQDVGIHARH----  - ISRV 97
                          :  :  :  :  | :: :  |  |          |
Db         85 GSPRISGDEEDDGDVSDFEMEFQIFNKSQDSHPQHENEEYNNNNHQWHPNGQAFSVAGST 144

Qy         98 STVDSEMNDE YAN  WKDRVKSRFGKKGKKKKSPFAETEPAQVPTEQQMEEKPSAEAS 156
              :  | :  :  |  :  |  :  :  :  :  :  :  :  |  :  :  |  :
Db        145 AGKDLEGDKEIVSM WKERVEKWKMWQLL  EKPGLVANDNGGNDPPEE--DDYLLAEAR 199

Qy        157 EPLSIVVPIPRNKLTPYRAVIIMRIVILGLFFHFRITNPVDSAFGLWLTSVICEIWFAFS 216
             :||    |  :  :::|  |::  ||  '  |:  |   :  ||| :||||:'|||
Db        200 QPLWRKVPISSSLISPYRIVIVLRFFILAFFLRFRILTPAYDAYPLWLISVICEVWFAFS 259

Qy        217 WVLDQFPKWNPVNRETYIERLSARYEREGEPNQLAQVDFFVSTVDPLKEPPLITANTVLS 276
             :|||||||  :  |  :|:::  :  ||||  |  :  |||||  ||||:'|||  '.'
Db        260 WILDQFPKWFPITRETYLDRLSLRYEREGEPNQLGPVDVFVSTVDLLKEPPLITANAVLS 319

Qy        277 ILAVDYPVDKVSCYVSDDGAAMLSFFSLVETAEFARKWVPFCKKFSIEPRAPEFYFSQKI 336
             ;|||||:|:|| |||:.:|||:' |:;: ||||||:|:|||||| ::|||||:'|:|:||
Db        320 ILAVDYPVEKVSCYVSDDGASMLLFDSLSETAEFARRWVPFCKKHNVEPRAPEFYFNEKI 379

Qy        337 DYLKDKVQPSFVKERRAMKREYEEFKVRVNALVAKAQKTPEEGWTMQDGTPWPGNNTRDH 396
             |||||| :|||:.'||||':|:|:|||:.||||:|| ||||| |||:|||||:||||
Db        380 DYLKDKVHPSFVKERRAMKREYEEFKVRINALVAKAQKKPEEGWVMQDGTPWPGNNTRDH 439

Qy        397 PGHDSGLPWEILGAFDIEGNELPRLVYVSREKRPGYQHHKKAGAENALVRVSAVLTNAPY 456
             ||    :    |:  ::|  ||| : |||'|||||:|||||:|||||:|||||| :
Db        440 PGMIQ VYLGSAGAL VDGKELPRLVYVSREKRPGYQHHKKAGAENALVRVSAVLTNAPF 498

Qy        457 ILNVDCDHYVNNSKAVREAMCILMDPQVGRDVCYVQFPQRFDGIDKSDRYANRNVVFFDV 516
             :|||:|||||:|'|||:'|||| :|||| : :'|||||||||||: '|||:'|||:::
Db        499 ILNLDCDHYINNSKAMREAMCFLMDPQFGKKLCYVQFPQRFDGIDRHDRYANRNVVFFDI 558

Qy        517 NMKGLDGIQGPVYVGTGCVFNRQALYGYGPPSMPSLRKRKDSSSCFS- --SCCCPSKKKP 572
             || :|||:.|||.'|||||||||||.||| :|   ' :: | :  |:     ||,   :|
Db        559 NMLGLDGLQGPVYVGTGCVFNRQALYGYDPP- VSEKRPKMTCDCWPSWCCCCGGSRKK 616

Qy        573 AQDEAE- ----   VYRDAKR -      EDLNAAIFNLTEI----DNYDEHERSMLISQL 623
             ::   /          ::     :  :|  .,:  ||     :  :.  |:  :|:|
Db        617 SKKKGEKKGLLGSL YGKKKKMMGKNYVKKGSAPVFILEEIEEGLEGYEEELEKSTLMSQK 676

Qy        624 SFEKTFGLSSVFIESTLMENGGVPESANSPPFIKEAIQVIGCGYEEKTEWGKQIGWIYGS 683
             :||| || | |||  :||..||:     |    ||: | || :|||| ||||: ||||.:
Db        677 NFEKRFGQSPVFIASTLMENGGLPESTNSTSLIKEAIHVISCGYEEKTEWGKEIGWIYGS 736

Qy        684 VTEDILSGFKMHCRGWRSIYCMPVRPAFKGSAPINLSDRLHQVLRWALGSVEIFFSRHCP 743
             |||||:|||:|||:.:|:|.:| |.:'||||.||||':|||||||||||:||| ||||
Db        737 VTEDILTGFKMHCRGWKSVYCVPKRPAFKGSAPINLSDRLHQVLRWALGSVEIFLSRHCP 796

Qy        744 LWYGFGGGRLKWLQRLAYINTIVYPFTSLPLIAYCTIPAVCLLTGKFIIPTLSNLASMLF 803
             ||||: ||:|||.:|||:|||||:|:|:|:|||||||||||||||'| |: |
Db        797 LWYGY-GGKLKWLERLAYINTIVYPFTSIPLLAYCTIPAVCLLTGKFIIPTLSNLTSVWF 855

Qy        804 LGLFISIIVTAVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQGFLKMLAGIDTNFT 863
```

```
Db        856 LALPLSIIATGVLELRWSGVSIQDWWRNEQFWVIGGVSAHLFAVFQGLLKVLAGVDTNFT 915

Qy        854 VTAKAAEDAEFGELYMVKWTTLLIPPTTLIIINMSG-CAGFSDALNKGYEAWGPLFGKVF 912
Db        916 VTAKAADDTEFGELYLFKWTTLLIPPTTLIILNMVGVVAGVSDAINNGYGSWGPLFGKLF 975

Qy        913 FAFWVILHLYPFLKGLMGRQNLTPTIVVLWSVLLASVFSLVWVKINPFVNKVDNTLVAET 972
Db        976 FAFWVILHLYPFLKGLMGRQNRTPTIVVLWSILLASIFSLVWVRIDPFLPKQTGPVLKQ- 1034

Qy        973 CISIDC 978
Db       1035 CGVEC 1039


RESULT 7
ABB93701
ID   ABB93701 standard; Protein; 1043 AA.
XX
AC   ABB93701;
XX
DT   31-MAY-2002  (first entry)
XX
DE   Herbicidally active polypeptide SEQ ID NO 2912.
XX
KW   Herbicidal; plant; agriculture; herbicide.
XX
OS   Arabidopsis thaliana.
XX
PN   WO200210210-A2.
XX
PD   07-FEB-2002.
XX
PF   28-AUG-2001; 2001WO-EP09892
XX
PR   29-AUG-2001; 2001WO-EP09892
XX
PA   (FARB ) BAYER AG.
XX
PI   Tietjen K, Weidler M;
XX
DR   WPI; 2002-269010/31.
XX
PT   Identifying plant target proteins for herbicidally active compounds,
PT   comprising aligning and comparing nucleic acid or amino acid sequences
PT   from plant with nucleic acid or amino acid sequences from non-plant
PT   organisms
XX
PS   Claim 5; SEQ ID NO 2912; 261pp + Sequence Listing; English.
XX
CC   The invention relates to identifying target proteins
CC   (ABB90790-ABB94016) for herbicidally active compounds, comprising
CC   aligning and comparing nucleic acid or amino acid sequences from plant
CC   with nucleic acid or amino acid sequences from non plant organisms using
CC   suitable search parameters, where plant sequences having an E-value
CC   greater by a factor of 3 than the E value of most similar non-plant
CC   sequences are selected. The polypeptides or nucleic acids encoding them
CC   are useful for identifying modulators. The identified modulators are
CC   useful as herbicides.
XX
SQ   Sequence  1043 AA;

  Query Match             66.2%;  Score 3470;  DB 21;  Length 1043;
  Best Local Similarity   63.6%;  Pred. No. 0;
  Matches  665;  Conservative  122;  Mismatches  170;  Indels   88;  Gaps   17;

Qy          6 APICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEINEGKKVCLRCGSPYDEN--  61
Db         15 AKICKVCGDEVKDDDNGQTFVACHVCVYPVCKPCYEYERSNGNKCCPQCNTLYKRHKGSP  74

Qy         62 -LLDDVEKKGSGN-----  --QSTMASHLN---  -  DEQDVGIHARHISSVSTV 100
```

```
Db         75 KIAGDEENNGPDDSDDELNIKYRQDGSSIHQNFAYGSVLFDFDKQQWRPNGRAFSSTGSV 134

Qy        101 ---DSEM-NDEYGNPIWKNRVKSCKDKENKKKKRSPKAETEPAQVPTEQQMEEKPSAEAS 156
Db        135 LGKDFEAERDGYTDAEWKERVDKWKARQ---EKRGLVTKGEQTNEDKEDDEEEYLDAEAR 191

Qy        157 EPLSIVYPIPRNKLTPYRAVIIMRLVILGLFFHFRITNPVDSAFGLWLTSVICEIWFAFS 216
Db        192 QPLWRKVPISSSKISPYRIVIVLRLVILVFFFRFRILTPAKDAYPLWLISVICEIWFALS 251

Qy        217 WVLDQFPKWNPVNRETYIERLSARYEREGEPSQLAGVDFFVSTVDPLKEPPLITANTVLS 276
Db        252 WILDQFPKWFPINRETYLDRLSMRFERDGEKNKLAPVDVFVSTVDPLKEPPIITANTILS 311

Qy        277 ILAVDYPVDKVSCYVSDDGAAMLSFESLVETAEFARKWVPFCKKFSIEPRAPEFYFSQKI 336
Db        312 ILAVDYPVNKVSCYVSDDGASMLLFDTLSETSEFARRWVPFCKKYNVEPRAPEFYFSEKI 371

Qy        337 DYLKDKVQPSFVKERRAMKRDYEEYKVRVNALVAKAQKTPEEGWTMQDGTPWPGNNTRDH 396
Db        372 DYLKDKVQTTFVKDRRAMKREYEEFKVRINALVAKAQKKPEEGWVMQDGTPWPGNNTRDH 431

Qy        397 PGHDSGLPWEIL-------GARDIEGNELPRLVYVSREKRPGYQHHKKAGAENALVRVSA 449
Db        432 PG--------MIQVYLGKEGAFDIDGNELPRLVYVSREKRPGYAHHKKAGAMNAMVRVSA 483

Qy        450 VLTNAPYILNVDCDHYVNNSKAVREAMCILMDPQVGRDVCYVQFPQRFDGIDKSDRYANR 509
Db        484 VLTNAPFMLNLDCDHYINNSKAIRESMCFLMDPQLGKKLCYVQFPQRFDGIDLNDRYANR 543

Qy        510 NVVFFDVNMKGLDGIQGPVYVGTGCVFNRQALYGYGPPSMPSLRKRKDSSSCF----SCC 565
Db        544 NIVFFDINMRGLDGIQGPVYVGTGCVFNRPALYGYEPP--VSEKRKKMTCDCWPSWICCC 601

Qy        566 C----------PSKKKPA--------------QDPAEVYRDAKREDLNAAIFNLTEI-- 598
Db        602 CGGGNRNHKSDSSKKKSGIKSLFSKLKKKTKKKSDDKTMSSYSRKRSSTEAIFDLEDIEE 661

Qy        599 --DNYDEHERSMLISQLSFEKTFGLSSVFIESTLMENGGVPESANSPPFIKEAIQVIGCG 656
Db        662 GLEGYDELEKSSLMSQKNFEKRFGMSPVFIASTLMENGGLPEATNTSSLIKEAIHVISCG 721

Qy        657 YEEKTEWGKQIGWIYGSVTEDILSGFKMHCRGWRSIYCMPVRPAFKGSAPINLSDRLHQV 716
Db        722 YEEKTEWGKEIGWIYGSVTEDILTGFRMHCRGWKSVYCMPKRPAFKGSAPINLSDRLHQV 781

Qy        717 LRWALGSVEIFFSRHCPLWYGFGGGRLKWLQRLAYINTIVYPFTSLPLIAYCTIPAVCLL 776
Db        782 LRWALGSVEIFFSRHCPLWYAW-GGKLKILERLAYINTIVYPFTSIPLLAYCTIPAVCLL 840

Qy        777 TGKFIIPTLSNLASMLFLGLFISIIVTAVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFA 836
Db        841 TGKFIIPTINNFASIWFLALFLSIIATAILELRWSGVSINDLWRNEQFWVIGGVSAHLFA 900

Qy        837 VFQGFLKMLAGIDTNFTVTAKAAEDA--EFGELYMVKWTTLLIPPTTLLIINMSG-CAGF 893
Db        901 VFQGLLKVLFGVDTNFTVTSKGASDEADEFGDLYLFKWTTLLIPPTTLIILNMVGVVAGV 960

Qy        894 SDALNKGYEAWGPLFGKVFFAFWVILHLYPFLKGLMGRQNLTPTIVVLWSVLLASVFSLV 953
Db        961 SDAINNGYGSWGPLFGKLFFAFWVIVHLYPFLKGLMGRQNRTPTIVVLWSILLASIFSLV 1020

Qy        954 WVKINPFVNKVDNTLVAETCISIDC 978
Db       1021 WVRIDPFLPKQTGPLLKQ--CGVDC 1043


RESULT 8
AAY58837
ID   AAY58837 standard; Protein; 1039 AA.
XX
AC   AAY58837;
XX
```

08-MAY-2000 (first entry)

Soybean cellulose synthase (full-length).

Soybean; cellulose synthase; transgenic plant.

Glycine max.

Key	Location/Qualifiers
Misc-difference	201
	/note= "encoded by AGAA (frameshift)"

WO200004166-A2.

27-JAN-2000.

13-JUL-1999; 99WO-US15871.

14-JUL-1998; 98US-0092844.

(DUPO) DU PONT DE NEMOURS & CO E I.

Allen SM, Fader GM, Falco SC, Kinney AJ, Lightner JE, Miao G, Rafalski JA, Thorpe CJ;

WPI; 2000-182431/16.
N-PSDB; AAZ58268.

Nucleic acid fragments encoding cellulose biosynthetic enzyme useful as probes for isolating cDNAs and genes encoding homologous proteins, for producing transgenic plants

Claim 6; Page 56-60; 93pp; English.

The present sequence is that of full-length soybean cellulose synthase (CS), as deduced from an isolated cDNA clone (see AAY58268). The invention relates to isolated CS nucleic acid fragments and to CS polypeptides. It also relates to the construction of a chimeric gene encoding all or a portion of the CS, in sense or antisense orientation, where expression of the gene results in altered levels of the CS in transformed host cells. The host cells can be used to screen compounds for their ability to inhibit CS activity. CS nucleic acids are also useful for producing transgenic plants having altered levels of CS, and hence altered levels of fibre. CS may also serve as a target for the development of novel herbicides.

Sequence 1039 AA;

Query Match 64.3%; Score 3367; DB 21; Length 1039;
Best Local Similarity 62.1%; Pred. No. 1.3e-309;
Matches 640; Conservative 118; Mismatches 182; Indels 90; Gaps 15;

```
Qy          8 ICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEINEGRKVCLRC--------GSPY- 58
              :|  ||: ||   :|:|||||:|| :|:|: |:|:|  ||  :| :|         |||
Db         36 VCEICGDGVGLTVDGDLFVACNECGFPVCRPCYEYERREGSHLCPQCKTRYKRLKGSPRV 95

Qy         59 ----DENLLDDVEKK----------------------GSGNQSTMASHLNDSQDVGIH 90
                ||  :||:| :                      | | :   |  :       :
Db         96 EGDDDEEDVDDIEHEFNIDEQKNKHGQVAEAMLHGRMSYGRGPEDDDNSQFPTPVIAGGR 155

Qy         91 ARHISSVSTVDSEMN-----------------DEYGNPIW-KNRVKSCKDKENKKKKRS 131
              :| :|    : |                   | |:  | : :    ||: :  | :
Db        156 SRPVSGEFPISSNAYGDQMLSSSLHKRVHPYPVSEPGSARWDEKKXDGWKDRMDDWKLQQ 215

Qy        132 PKAETEPAQVPTEQQMEEKPSAEASEPLSIVYPIPRNKLTPYRAVIIMRLVILGLFFHFR 191
                || : |    ::     || :|||   ||  :|: ||| ||: ||||  :  :|
Db        216 GNLGPEPDEDPDAAMLD-----EARQPLSRKVPIASSKINPYRMVIVARLVILAFFLRYR 270

Qy        192 ITNPVDSAFGLWLTSVICEIWFAFSWVLDQFPKWNPVNRETYIERLSARYEREGEPSQLA 251
              : |||  | ||||||:|||||||||:|||||| |::||||::||| ||||||||: ||
Db        271 LMNPVHDALGLWLTSIICEIWFAFSWILDQFPKWFPIDRETYLDRLSIRYEREGEPNMLA 330

Qy        252 GVDFFVSTVDPLKEPPLITANTVLSILAVDYPVDKVSCYVSDDGAAMLSFESLVETAEFA 311
```

```
Db        331 PVDVFVSTVDPMKEPPLVTANTVLSILAVDYPVDKISCYISDDGASMLTFESLSETAEFA 390

Qy        312 RKWVPFCKKFSIEPRAPEFYFSQKIDYLKDKVQPSFVKERRAMKRDYEEYKVRVNALVAK 371
              |||:|||:|||:: ||| :||: :-|||:  :||: : ||||.|:|:|||::||:|||||
Db        391 RKWVPFCKKFSIEPRAPEMYFSRKIDYLKDKVQPTFVKERRAMKREYEEFKVRINALVAK 450

Qy        372 AQKTPEEGWTMQDGTPWPGNNTRDHPGMIQVFLGSSGGLDTEGNQLPRLVYVSREK 428
              ||| |: :| |' |||:|:|||| ||    |:    ||:    | |::||||:|'||||
Db        451 AQKVPQGGWIMQDGTPWPGNNTKDHP-   GMIQVFLGSSGGLDTEGNQLPRLVYVSREK 506

Qy        429 RPGYQHHKKAGAENALVRVSAVLTNAPYILNVDCDHYVNNSKAVREAMCILMDPQVGRDV 488
              |||:|||||||| |.' ||| || |:::||: :||||:|||| |||:| |||| |: :
Db        507 RPGFQHHKKAGAMNALVRVSAVLTNAPYMLNLDCDHYVNNSKAAREAMCFLMDPQTGKRV 566

Qy        489 CYVQFPQRFDGIDK- RYANRNVVFFDVNMKGLDGIQGPVYVGTGCVFNRQALYGYGPPS 548
              ||||:|||:||| '||:'| |' |:|:'||||:|||||||||| |||| |||:| |: :
Db        567 CYVQFPQRFDGIDTH RYANRNTVFFDINMKGLDGIQGPVYVGTGCVFRRQALYGYNPPK 626

Qy        549 MPSLRKRKDSSS-P  CFPKFFRQR AEVYRDAFREDLNAAIFNLTEIDNYDEHERS 607
              |  ||    |.     |:|       |::   :| :   :|  :|:    ::
Db        627 GP-- KRPKMVS-- ---------FFF  KNDANGEAASLKGMDL -- GKE 669

Qy        608 MLISQDLSFEKTF-  -FRMSGVPVSGS AALLKEATHVISCGYEDKTEWGLEL 662
              :|:||:: || :  :|:  :|| |  :|  :|:|:| || ::||:|  || ::
Db        670 VLMSQMNFEKKFG-  -FPVTSTLMEKG IVPPSGS AALLKEATHVISCGYEDKTEWGLEL 729

Qy        668 GWIYGSVTEDILTGFKMHCRGWRSIYCMPVRPAFKGSAPINLSDRLHQVLRWALGSVEIF 727
              ||||:.|||||:|.|'|||||||||:'| | :||:|:|||||||:|||||.|||:|.
Db        730 GWIYGSITEDILTGFKMHCRGWRSIYCMPKRAAFKGTAPINLSDRLNQVLRWALGSIEIF 789

Qy        728 FSRHCPLWYGFGGRLKWLQRLAYINTIVYPFTSLFLIAYCTIPAVCLLTGKFIIPTLSN 787
              || ||||||!|  |||:|:| |. :| |:|||:|:|||  :|||||| !||:| :'
Db        790 FSHHCPLWYGFKGKLKWLERFAYANTVVYPFTSIPLVAYCILPAVCLLTDKFIMPPIST 849

Qy        788 LASMLFLGLFISIITATVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQGLLKMLAG 847
              |: : :| :||  :|||: |:|   : | | ||||:|||||':|| |  :|||:|:|||
Db        850 FAGLYFVALFSSIIATSILELKWGGVSIFEWWRNEQFWVIGGVSAHLFAVIQGLLKVLAG 909

Qy        848 IDTNFTVTAKAAEDAEFGELYMVKWTTLLIPPTTLLIINMSG-CAGFSDALNKGYEAWGP 906
              |||||||:': :| :|||||  || .'|||:-:'||||:: | ;; |||:' ||:::!
Db        910 IDTNFTVTSKATDDEEFGELYTFKWTTLLIPPTTLLIINIVGVVAGISDAINNGYQSWGP 969

Qy        907 LFGKVFFAFWVIMH YPFLKGLMGRQNRTPTIVVIWSVLLASVFSLVWVKINPFVNKV  964
              ||'.:|||:|||:  ||||'|||  :|| |  |||:  '||||||||:|:|:|:|||
Db        970 LFGKLFFSFWVIVH YPFLKGLMGRQNRTPTIVVIWSVLLASIFSLLWVRIDPFVLKTKG 1029

Qy        965 -DNTLVAETC 973
               |  |  |
Db       1030 PDTKLCGINC 1039
```

RESULT 9
AAU10496
ID AAU10496 standard; Protein; 1079 AA
XX
AC AAU10496;
XX
DT 14-FEB 2002 (first entry)
XX
DE Corn cellulose synthase Cqrael9/cesA 9.
XX
KW Corn; cellulose synthase; Cdpgs45; cesA-3; Cqrael9; cesA-9;
KW stalk quality; improved stand; silage; pericarp; kernel hardening;
KW handling ability; transgenic plant; immunogen.
XX
OS Zea mays.
XX
PN WO200179516-A2.
XX
PD 25-OCT-2001.
XX
PF 12-APR-2001; 2001WO-US11951.

```
XX
PR   14 APR 2000; 2000US-0550483.
XX
PA   (PION-) PIONEER HI-BRED INT INC.
XX
PI   Dhugga KS, Helentjaris TG;
XX
DR   WPI; 2002-041338/05.
DR   N-PSDB; AAS16458.
XX
PT   New cellulose synthase polypeptides and polynucleotides, useful in
PT   improving stalk quality or silage, and in increasing concentration of
PT   cellulose in the pericarp, hardening the kernel for improved handling
PT   ability
XX
PS   Claim 3; Page 85-87; 98pp; English.
XX
CC   The invention relates to isolated nucleic acids encoding two cellulose
CC   synthase proteins from corn, Cdpgs45 (cesA-3) and Cgrael9 (cesA-9).
CC   Also disclosed are a recombinant expression cassette comprising the
CC   polynucleotide (operably linked to a promoter) a host cell comprising
CC   the recombinant expression cassette and a transgenic plant comprising
CC   the recombinant expression cassette. The nucleic acid is useful in
CC   the improvement of stalk quality for improved stand or silage, and in
CC   the increased concentration of cellulose in the pericarp, hardening
CC   the kernel, and thus improving its handling ability. The nucleic acids
CC   may also be used as probes or amplification primers in the detection,
CC   quantification or isolation of gene transcripts, as probes in detecting
CC   deficiencies in the level of mRNA, for detecting gene mutations or
CC   allelic variants, for monitoring up regulation of expression or changes
CC   in enzyme activity in screening assays, for site directed mutagenesis,
CC   and in sense or antisense suppression of one or more genes in a host
CC   cell, tissue or plant. The polypeptides may be used in assays for
CC   enzyme agonists or antagonists, and as immunogen or antigen to
CC   obtain antibodies specifically immunoreactive with the protein.
CC   The present sequence represents a corn cellulose synthase of the
CC   invention.
XX
SQ   Sequence   1079 AA;

  Query Match             64.2%;  Score 3366.5;  DB 23;  Length 1079;
  Best Local Similarity   59.6%;  Pred. No. 1.6e-309;
  Matches  645;  Conservative  133;  Mismatches  179;  Indels  127;  Gaps   19;

Qy          5 GAPLTHTCGEQVGHDANGELPVATRRQVFMGRSPEFPEPIDSGRKVCLRC-------GS 56
              |  :   ||: :|  | |: : |  :  |  :  :|:|   |  :   :|        ||
Db         15 GGQVCQICGDDVGTTAAGDVFTACVICLPVCRICYEYERE-NEQACPQCKNKYKRHKGS 74

Qy         57 P- ------  -  -  -  YDENDEEVDERKSDKQSTMVSELNDSQDVG -------- 88
              |                  |  :  :|  ::  :   :   :  ' |||
Db         75 PAIRGEEGDDTDADDASDFNYPASGNDDQKQKIADRMRSWRMNAGGSGDVGRPKYDSGEI 134

Qy         89 ---------IHARHISSV -------              -  ------    ---STVDS 102
                       |  :| ||                                    |
Db        135 GLTKYDSGEIPRGYIPSVTNSQISGEIPGASPDHHMMSPTGNIGRRAPFPYMNHSSNPSR 194

Qy        103 EMNDEYGNPIWKNRVKSCKDKENK -  KKKRSPKAETEPAQV  ---PTEQQMEEK-PS 152
              |  :   |,  || ||   ,|:::      :  | :|   |       |:  ||:   :
Db        195 EPSGSVGNVAWKERVDGWKMKQDKGTIPMTNGTSIAPGEGLGVGDIDASTDYNMEDALLN 254

Qy        153 AEASEPLSIVYPIPRNKLTPYRAVIIMRLVILGLFFHFRITNPVDSAFGLWLTSVICEIW 212
                 |  :|||   |:| ::: ||| '|:::'::| :: |:|: ||| :|: ||| |||||||
Db        255 DETRQPLSRKVPLPSSRINPYRMVIVLRLIVLSIFLHYRITNPVRNAYPLWLLSVICEIW 314

Qy        213 FAFSWVLDQFPKWNPVNRETYIERLSARYEREGEPSQLAQVDIFVSTVDPLKEPPLITAN 272
              || ||:|-|||||  ::|||| ::|:: ||  ||||||: | |  |||:|||:||||||:||
Db        315 FALSWILDQFPKWFPINRETYLDRLALRYDREGEPSQLAAVDIFVSTVDPMKEPPLVTAN 374

Qy        273 TVLSILAVDYPVDKVSCYVSDDGAAMLSFESLVETAEFARKWVPFCKKFSIEPRAPEFYF 332
              |'||||||'||||||-||||-|||| ::|::' ||:||'|||: | |||:|||||||:||
Db        375 TVLSILAVDYPVDKVSCYVSDDGAAMLTFDALAETSEFARKWVPFVKKYNIEPRAPEWYF 434

Qy        333 SQKIDYLKDKVQPSFVKERRAMKREYEEFKVRVNALVAKAQKTPEEGWTMQDGTPWPGNN 392
              |||||||||| |||||:|||||:|||:|:||| ||||||| ||||| ||||||||||
Db        435 SQKIDYLKDKVHPSFVKDRRAMKREYEEFKIRVNGLVAKAQKVPEEGWIMQDGTPWPGNN 494

Qy        393 TRDHP-------GHDSGLPWEILGARDIEGNELPRLVYVSREKRPGYQHHKKAGAENALV 445
              ||||        ||  ||         | ||||||||||||||||:|||||||| ||||
Db        495 TRDHPGMIQVFLGHSGGL--------DTEGNELPRLVYVSREKRPGFQHHKKAGAMNALV 546

Qy        446 RVSAVLTNAPYILNVDCDHYVNNSKAVREAMCILMDPQVGRDVCYVQFPQRFDGIDKSDR 505
              ||||||||  |:||:|||||:|||||:|||| :|| ||||||||||||||::||
Db        547 RVSAVLTNGQYMLNLDCDHYINNSKALREAMCFLMDPNLGRSVCYVQFPQRFDGIDRNDR 606

Qy        506 YANRNVVFFDVNMKGLDGIQGPVYVGTGCVFNRQALYGYGPPSMPSLRKRKDSSSCFSCC 565
              ||||| ||||:|::|||||||||||||||||||| ||||| ||    ::::|      |
Db        607 YANRNTVFFDINLRGLDGIQGPVYVGTGCVFNRTALYGYEPP----IKQKK--GGFLSSL 660

Qy        566 CPSKKKPAQDP--AEVYRDAKREDLNAAIFNLTEID------NYDEHERSMLISQLSFEK 617
              |  :|| ::    ::  :  |  | :  :||| :|:       :|: |:|:|:||:| ||
Db        661 CGGRKKGSKSKKGSDKKKSQKHVDSSVPVFNLEDIEEGVEGAGFDD-EKSLLMSQMSLEK 719

Qy        618 TFGLSSVFIESTLMENGGVPESANSPPFIKEAIQVIGCGYEEKTEWGKQIGWIYGSVTED 677
               || |: |: ||||| || |:||     :|||| || ||||:| ||| :||||||||||
Db        720 RFGQSAAFVASTLMEYGGVPQSATPESLLKEAIHVISCGYEDKIEWGTEIGWIYGSVTED 779

Qy        678 ILSGFKMHCRGWRSIYCMPVRPAFKGSAPINLSDRLHQVLRWALGSVEIFFSRHCPLWYG 737
              ||:|||| ||||||||| ||||||||||||||:|||||||||| |||||'|||
Db        780 ILTGFKMHARGWRSIYCMPKRPAFKGSAPINLSDRLNQVLRWALGSVEILFSRHCPLWYG 839

Qy        738 FGGGRLKWLQRLAYINTIVYPFTSLPLIAYCTIPAVCLLTGKFIIPTLSNLASMLFLGLF 797
              : ||||:|:| |||||  :|| |||||: || :||:|||||||||| :|| ||: |: ||
Db        840 Y-GGRLKFLERFAYINTTIYPLTSLPLLIYCILPAICLLTGKFIIPEISNFASIWFISLF 898

Qy        798 ISIIVTAVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQGFLKMLAGIDTNFTVTAK 857
              |||  | :||:|||||  |:: ||||||||||:|||||||||  ||:|||||||||:|
Db        899 ISIFATGILEMRWSGVGIDEWWRNEQFWVIGGISAHLFAVFQGLLKVLAGIDTNFTVTSK 958

Qy        858 AA-EDAEFGELYMVKWTTLLIPPTTLLIINMSG-CAGFSDALNKGYEAWGPLFGKVFFAF 915
              |: || :| |||| ||||||||||:|||: |  || | |:| ||::|||||||:|||!
Db        959 ASDEDGDFAELYMFKWTTLLIPPTTILIINLVGVVAGISYAINSGYQSWGPLFGKLFFAF 1018

Qy        916 WVILHLYPFLKGLMGRQNRTPTIVVLWSVLLASVFSLVWVKINPFVNKVDNTLVAETCIS 975
              |||:|||||||||||||| ||||||:|::||||:|||:||:|:|| |:|       ||
Db       1019 WVIVHLYPFLKGLMGRQNRTPTIVVVWAILLASIFSLLWVRIDPFTNRVTGP-DTRTC-G 1076

Qy        976 IDC 978
              |:|
Db       1077 INC 1079


RESULT 10
AAY84108
ID   AAY84108 standard; Protein; 1077 AA.
XX
AC   AAY84108;
XX
DT   03-JUL-2000  (first entry)
XX
DE   Amino acid sequence of a maize cellulose synthase.
XX
KW   Maize; cellulose synthase; stalk quality; stand; silage; cellulose;
KW   transgenic plant; plant breeding marker.
XX
OS   Zea mays.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 494
FT                   /note= "encoded by MGG"
XX
PN   WO200009706-A2.
XX
PD   24-FEB-2000.
XX
PF   16-AUG-1999;   99WO-US18760.
XX
```

17-AUG-1998; 98US-0096822.

(PION-) PIONEER HI-BRED INT INC.

Dhugga KS, Helentjaris TG, Bowen BA, Wang X;

WPI; 2000-224343/19.
N-PSDB; AAZ99494.

New genes which encode maize cellulose synthase polypeptides in plants useful for modulating the expression of cellulose synthase in plants and to produce transgenic plants expressing the novel protein -

Claim 15; Page 95-97; 119pp; English.

The present sequence represents a maize cellulose synthase polypeptide. The cellulose synthase can be used for the improvement of stalk quality for improved stand or silage. It also provides an increased concentration of cellulose in the pericarp, hardening the kernel and improving its handling ability. The sequences are used to produce transgenic plants and seeds expressing the cellulose synthase. The polynucleotide is used for modulating, preferably increasing, the level of the synthase in a plant cell. The plants are preferably monocots. The polynucleotide is also used as a probe or primer in the detection quantitation or isolation of gene transcripts. The probes are useful in detecting deficiencies in the level of mRNA in screenings for desired transgenic plant, for detecting mutations in the gene, for monitoring upregulation of expression or changes in enzyme activity in screening assays of compounds, for detection of any number of allelic variants of the gene, or for use as molecular markers in plant breeding programs. The isolated nucleic acids of the present invention can also be used for recombinant expression of their encoded polypeptides or for use as immunogens in the preparation and/or screening of antibodies. The proteins can be employed in assays for enzyme agonists or antagonists of enzyme function or for use of immunogens or antigens to obtain antibodies specifically immunoreactive with a protein.

Sequence 1077 AA;

Query Match 64.1%; Score 3359.5; DB 21; Length 1077;
Best Local Similarity 59.6%; Pred. No. 7.1e-309;
Matches 644; Conservative 136; Mismatches 176; Indels 125; Gaps 20;

```
Qy          5 GAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEINEGRKVCLRC--------GS 56
              |  :|  ||: ||  | |::| ||  | :|:|: |:|:|  :| : | :|        ||
Db         15 GGQVCQICGDGVGTTAEGDVFAACDVCGFPVCRPCYEYERKDGTQACPQCKTKYKRHKGS 74

Qy         57 P------------YDENLL----DDVEKKGSGNQSTMASHLNDSQDVG---------- 88
              |            | | |    :| ::| :    :   ::  | |||
Db         75 PAIRGEEGDDTDADSDFNYLASGNEDQKQKIADRMRSWRMNVGGSGDVGRPKYDSGEIGL 134

Qy         89 -------IHARHISSVS------------------ ------------------TVDSEM 104
                    |   :| ||:                                       |
Db        135 TKYDSGEIPRGYIPSVTNSQISGEIPGASPDHHMMSPTGNIGKRAPFPYVNHSPNPSREF 194

Qy        105 NDEYGNPIWKNRVKSCKDKENK----KKKRSPKAETEPAQV-----PTEQQMEEK-PSAE 154
              :   ||  || ||   | |::|        :  | :|   |      |:  ||:   : |
Db        195 SGSIGNVAWKERVDGWKMKQDKGTIPMTNGTSIAPSEGRGVGDIDASTDYNMEDALLNDE 254

Qy        155 ASEPLSIVYPIPRNKLTPYRAVIIMRLVILGLFFHFRITNPVDSAFGLWLTSVICEIWFA 214
                :|||   |:| ::: ||| ||::||::| :| |:|||||| :|: ||| ||||||||
Db        255 TRQPLSRKVPLPSSRINPYRMVIVLRLIVLSIFLHYRITNPVRNAYPLWLLSVICEIWFA 314

Qy        215 FSWVLDQFPKWNPVNRETYIERLSARYEREGEPSQLAGVDFFVSTVDPLKEPPLITANTV 274
              ||:|||||| |:|||||:||: ||:|||||||| || |||||||:|||||:|||||
Db        315 LSWILDQFPKWFPINRETYLDRLALRYDREGEPSQLAAVDIFVSTVDPMKEPPLVTANTV 374

Qy        275 LSILAVDYPVDKVSCYVSDDGAAMLSFESLVETAEFARKWVPFCKKFSIEPRAPEFYFSQ 334
              |||||||||||||||||||||||||:|::| ||:|||||||| ||::|||||:||||
Db        375 LSILAVDYPVDKVSCYVSDDGAAMLTFDALAETSEFARKWVPFVKKYNIEPRAPEWYFSQ 434

Qy        335 KIDYLKDKVQPSFVKERRAMKRDYEEYKVRVNALVAKAQKTPEEGWTMQDGTPWPGNNTR 394
              |||||||| |||||:||||||:|||:||||| ||||||| ||||| ||||||||||||
```

```
Db        435 KIDYLKDKVHPSFVEERAMKREYEEFKVRVNGLVAKAQKVPEEGWIMQDGTPWPGNNTX 494

Qy        395 DHP ------GHDSGLPWEILGAFDIEGNELPRLVYVSREKRPGYQHHKKAGAENALVRV 447
              ||;       |  ||        | |:: |||:| ||||:|:|:||:|||| :|||
Db        495 DHPGMIQVFLGHSGGL------ DTEGNELPRLVYVSREKRPGFQHHKKAGAMNALVRV 546

Qy        448 SAVLTNAPYILNVDCDHYVNNSKAVREAMCFLMDPQVGRDVCYVQFPQRFDGIDKSDRYA 507
              ||||.|  |:||:|:|:||:|||:|||. ||| :.| ||||||||:|||:|||:||||
Db        547 SAVLTNGQYMLNLDCDHYINNSKALREAMCFLMDPNLGRSVCYVQFPQRFDGIDRNDRYA 606

Qy        508 NRNVVFFDVNMKGLDGIQGPVYVGTGCVFNRQALYGYGPPSMPSLRKRKDSSSCFSCCCP 567
              ||| ||.|:|::|||:|:|||| ||| : || : || ||    ::::|      | |
Db        607 NRNTVFFDINLRGLDGIQGPVYVGTGCVFNRTALYGYEPP----IKQKK- GGFLSSLCG 660

Qy        568 SKKKPAQDP--AEVYEAKREDLNAAIFNLTEID    ---NYDEHERSMLISQLSFEKTF 619
              :|| ::    ::  :    |   : ' :|:      :|: |:|:|:|||: || :
Db        661 GRKKASKSKKGSPKKKNQKHVDSSVPVFNLEDIEEGVEGAGFDD-EKSLLMSQMSLEKRF 719

Qy        620 GLSSVFIESTLMENGGVPESANPQTLIKEAIQVISCGYEEKTEWGKQIGWIYGSVTEDIL 679
              | |: |: ||||  :|:|:|      :|  || :|:|||:|:||||:|||
Db        720 GQSAAFVASTLMEY PDSATPPS PPAAIDIAVYEDKTEWGTEIGWIYGSVTEDIL 779

Qy        680 SGFKMHCRGWRSIY MYPARPAFKGSAPINLSDRLNQVLRWALGSVEIFFSRHCPLWYGFG 739
              :|||| |||:  ||  |||  :|   :|  :||||:|||||:|||||  ||||:
Db        780 TGFKMHARGWPSIY MEPARPAFKGSAPINLSDRLNQVLRWALGSVEILFSRHCPLWYGY- 838

Qy        740 GGRLKWLQRLAYINTIVYPFTSLPLIAYCTIPAVCLLTGKFIIPTLSNLASMLFLGLFIS 799
              ||||:|:  || :| :: ||:  : |, :||:: :|||||:| :|| ||: |: :|||
Db        839 GGRLKFLERFAYINTTIYPLTSIPLLIYCILPAICLLTGKFIIPEISNFASIWFISLFIS 898

Qy        800 IIVTAVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQGFLKMLAGIDTNFTVTAKAA 859
              |  : :||:|||||  :: ||||:|:|||: |||||:|| ||:|||||:|:|||:|:|:
Db        899 IFATGILEMRWSGVGIDEWWRNEQFWVIGGISAHLFAVFQGLLKVLAGIDTNFTVTSKAS 958

Qy        860 -EDAEFGELYMVKWTTLLIPPTTLLIINMSG CAGFSDALNKGYEAWGPLFGKVFFAFWV 917
               || :| |||| |||.:'|||: :|| :: |  |  | |:: ||::| |||':|||:'
Db        959 DEDGDFAELYMFKWTTLLIPPTTILIINLVGVVAGISYAINSGYQSWGPLFGKIFFAFWV 1018

Qy        918 ILHLYPFLKGLMGRQNRTPTIVVLWSVLLASVFSLVWVKINPFVNKVDNTLVAETCISID 977
              |:|||||:|||: || :|||||:::|||':|| : |:|::'  :|       :|,  |:
Db       1019 IVHLYPFLKGLMGRQNRTPTIVVVWAILLASIFSLLWVRIDPFTTRVTGP-DTQTC GIN 1076

Qy        978 C 978
              :
Db       1077 C 1077
```

```
RESULT 11
AAY84114
ID   AAY84114 standard; Protein; 1077 AA.
XX
AC   AAY84114;
XX
DT   03-JUL-2000 (first entry)
XX
DE   Amino acid sequence of a maize cellulose synthase.
XX
KW   Maize; cellulose synthase; stalk quality; stand; silage; cellulose;
KW   transgenic plant; plant breeding marker.
XX
OS   Zea mays.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 494
FT                   /note "encoded by MQ5"
XX
PN   WO200009706-A2.
XX
PD   24-FEB-2000.
XX
PF   16-AUG-1999;   99WO-US18760.
XX
PR   17-AUG-1998;   98US-0096822.
```

```
XX
PA   (PION ) PIONEER HI-BRED INT INC.
XX
PI   Dhugga KS, Helentjaris TG, Bowen BA, Wang X;
XX
DR   WPI; 2000-224343/19.
DR   N-PSDB; AAZ39512.
XX
PT   New genes which encode maize cellulose synthase polypeptides in plants
PT   useful for modulating the expression of cellulose synthase in plants
PT   and to produce transgenic plants expressing the novel protein -
XX
PS   Claim 15; Page 142-144; 119pp; English.
XX
CC   The present sequence represents a maize cellulose synthase polypeptide.
CC   The cellulose synthase can be used for the improvement of stalk quality
CC   for improved stand or silage. It also provides an increased concentration
CC   of cellulose in the pericarp, hardening the kernel and improving its
CC   handling ability. The sequences are used to produce transgenic plants
CC   and seeds expressing the cellulose synthase. The polynucleotide is
CC   used for modulating, preferably increasing, the level of the synthase
CC   in a plant cell. The plants are preferably monocots. The polynucleotide
CC   is also used as a probe or primer in the detection quantitation or
CC   isolation of gene transcripts. The probes are useful in detecting
CC   deficiencies in the level of mRNA in screenings for desired transgenic
CC   plant, for detecting mutations in the gene, for monitoring upregulation
CC   of expression or changes in enzyme activity in screening assays of
CC   compounds, for detection of any number of allelic variants of the gene,
CC   or for use as molecular markers in plant breeding programs. The
CC   isolated nucleic acids of the present invention can also be used for
CC   recombinant expression of their encoded polypeptides or for use as
CC   immunogens in the preparation and/or screening of antibodies. The
CC   proteins can be employed in assays for enzyme agonists or antagonists
CC   of enzyme function or for use of immunogens or antigens to obtain
CC   antibodies specifically immunoreactive with a protein.
XX
SQ   Sequence   1077 AA;

  Query Match             64.1%;  Score 3359.5;  DB 2;  Length 1077;
  Best Local Similarity   59.6%;  Pred. No. 7.1e-309;
  Matches  644;  Conservative  136;  Mismatches  176;  Indels  125;  Gaps   20;

Qy          5 GAPICHTCGEQVGHDANGELFVACHECAYEVCKSCFEFELKEGRKVCERC---  ---GS 56
              |  :  ||: |   |  : :| |  |  :  |: :|  :  : :|        ||
Db         19 GSQVCQICGDGVGTTAEGDVFAACDVCGFPVCRPCYEYERKEGTQACPQCKTKYKRHKGS 74

Qy         57 P   -- -------YDENDE ---EDVEKEDSGNQSTMASGLNLSQDVG  ----- -- 88
              |             | | '    :| :: |  :    :  :::  | |':
Db         75 PAIRGEEGDDTDADSDFNYLASGNEDQKQKIAERMRSWRMNVGGSGDVGRPKYDSGEIGL 134

Qy         89        -THARHISSVS-  -  ----- -- ---- -     - ----------TVDSEM 104
                     :  :| |::          :                          |
Db        135 TKYDSGEIPRGYIPSVTNSQISGEIPGASPDHHMMSPTGNIGKMAPFPYVNHSPNPSREF 194

Qy        105 NDEYGNPIWKNRVKSCKDKENK  --KKKPSFKAETEPAQV-  --PTEQQMEEK-PSAE 154
                :  ||  |! ||   | |::'       :  ' :|   '       |:  ||:   : |
Db        195 SGSIGNVAWKERVDGWKMKQDKGTIPMTNGTSIAESGRVVGDIDASTDYNMEDALLNDE 254

Qy        155 ASEPLSIVYPIPRNKLTPYRAVIIMRLVILGLFFHFRITNPVDSAFGLWLTSVICEIWFA 214
                 :'|'   |:| ::: :'| |::'|::  :| |:'||| : :|: ||| ||||||||
Db        255 TRQPLSRKVPLPSSRINPYRMVIVLRLIVLSIFLHYRITNPVRNAYPLWLLSVICEIWFA 314

Qy        215 FSWVLDQFPKWNPVNRETYIERLSARYERDGEPSQLAGVDIFVSTVDPLKEPPLITANTV 274
              |::|||\||| |:|'|||::.|: |::|||:|||| | |::||||:|||||:|||'|
Db        315 LSWILDQFPKWFPINRETYLDRLALRYDREGEPSQLAAVDIFVSTVDPMKEPPLVTANTV 374

Qy        275 LSILAVDYPVDKVSCYVSDDGAAMLSFESLVETAEFARKWVPFCKKFSIEPRAPEFYFSQ 334
              |:|||||||||||||||||| ||:|::| |::|||| |||  ||::|||||:||||
Db        375 LSILAVDYPVDKVSCYVSDDGAAMLTFDALAETSEFARKWVPFVKKYNIEPRAPEWYFSQ 434

Qy        335 KIDYLKDKVQPSFVKERRAMKREYEEYKVRVNALVAKAQKTPEEGWTMQDGTPWPGNNTR 394
              ||:|:|||| ||||:| |||:::|:|'|'| |||||: '|||| ||||'||||||
Db        435 KIDYLKDKVHPSFVKDRRAMKREYEEFKVRVNGLVAKAQKVPEEGWIMQDGTPWPGNNTX 494

Qy        395 DHP-------GHDSGLPWEILGARDIEGNELPRLVYVSREKRPGYQHHKKAGAENALVRV 447
              |||       ||  ||         | ||||||||||||||||:|||||||| ||||||
Db        495 DHPGMIQVFLGHSGGL--------DTEGNELPRLVYVSREKRPGFQHHKKAGAMNALVRV 546

Qy        448 SAVLTNAPYILNVDCDHYVNNSKAVREAMCILMDPQVGRDVCYVQFPQRFDGIDKSDRYA 507
              ||||||  |:||:|||||:|||||:|||| |||| :|| ||||||||||||::||||
Db        547 SAVLTNGQYMLNLDCDHYINNSKALREAMCFLMDPNLGRSVCYVQFPQRFDGIDRNDRYA 606

Qy        508 NRNVVFFDVNMKGLDGIQGPVYVGTGCVFNRQALYGYGPPSMPSLRKRKDSSSCFSCCCP 567
              ||| |||:|::|||||||||||||||||| ||||| ||    ::::|      |  |
Db        607 NRNTVFFDINLRGLDGIQGPVYVGTGCVFNRTALYGYEPP----IKQKK--GGFLSSLCG 660

Qy        568 SKKKPAQDP--AEVYRDAKREDLNAAIPNLTEID------NYDEHERSMLISQLSFEKTF 619
              :|| ::    ::  :  |  | :  :||| :|:         :|: |:|:|:||:| || |
Db        661 GRKKASKSKKGSDKKKSQKHVDSSVPVFNLEDIEEGVEGAGFDD-EKSLLMSQMSLEKRF 719

Qy        620 GLSSVFIESTLMENGGVPESANSPPFIKEAIQVIGCGYEEKTEWGKQIGWIYGSVTEDIL 679
              | |: |: ||||| ||||:||     :|||| || ||||:||||| :||||||||||||
Db        720 GQSAAFVASTLMEYGGVPQSATPESLLKEAIHVISCGYEDKTEWGTEIGWIYGSVTEDIL 779

Qy        680 SGFKMHCRGWRSIYCMPVRPAFKGSAPINLSDRLHQVLRWALGSVEIFFSRHCPLWYGFG 739
              :||||| |||||||||| ||||||||||||||||:|||||||||||| ||||||||||:
Db        780 TGFKMHARGWRSIYCMPKRPAFKGSAPINLSDRLNQVLRWALGSVEILFSRHCPLWYGY- 838

Qy        740 GGRLKWLQRLAYINTIVYPFTSLPLIAYCTIPAVCLLTGKFIIPTLSNLASMLFLGLFIS 799
              |||||:|:| ||||| :|| ||:||: || :||:|||||||||| :|| ||: |: ||||
Db        839 GGRLKFLERFAYINTTIYPLTSIPLLIYCILPAICLLTGKFIIPEISNFASIWFISLFIS 898

Qy        800 IIVTAVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQGFLKMLAGIDTNFTVTAKAA 859
              |  | :||:|||||  |:: ||||||||||:|||||||||| ||:|||||||||||:||:
Db        899 IFATGILEMRWSGVGIDEWWRNEQFWVIGGISAHLFAVFQGLLKVLAGIDTNFTVTSKAS 958

Qy        860 -EDAEFGELYMVKWTTLLIPPTTLLIINMSG-CAGFSDALNKGYEAWGPLFGKVFFAFWV 917
               || :| |||| ||||||||||:||||: |  || | |:| ||::|||||||:||||||
Db        959 DEDGDFAELYMFKWTTLLIPPTTILIINLVGVVAGISYAINSGYQSWGPLFGKLFFAFWV 1018

Qy        918 ILHLYPFLKGLMGRQNLTPTIVVLWSVLLASVFSLVWVKINPFVNKVDNTLVAETCISID 977
              |:|||||||||||| ||||||:|::||||:|||:||:|:||  :|      :||  |:
Db       1019 IVHLYPFLKGLMGRQNRTPTIVVVWAILLASIFSLLWVRIDPFTTRVTGP-DTQTC-GIN 1076

Qy        978 C 978
              :
Db       1077 C 1077


RESULT 12
AAY84121
ID   AAY84121 standard; Protein; 1076 AA.
XX
AC   AAY84121;
XX
DT   03-JUL-2000  (first entry)
XX
DE   DNA encoding a maize cellulose synthase.
XX
KW   Maize; cellulose synthase; stalk quality; stand; silage; cellulose;
KW   transgenic plant; plant breeding marker.
XX
OS   Zea mays.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 425
FT                   /note= "encoded by ANG"
XX
PN   WO200009706-A2.
XX
PD   24-FEB-2000.
XX
PF   16-AUG-1999;   99WO-US18760.
XX
PR   17-AUG-1998;   98US-0096822.
XX
```

(PION-) PIONEER HI-BRED INT INC.

Dhugga KS, Helentjaris TG, Bowen BA, Wang X;

WPI; 2000-224343/19.
N-PSDB; AAZ99533.

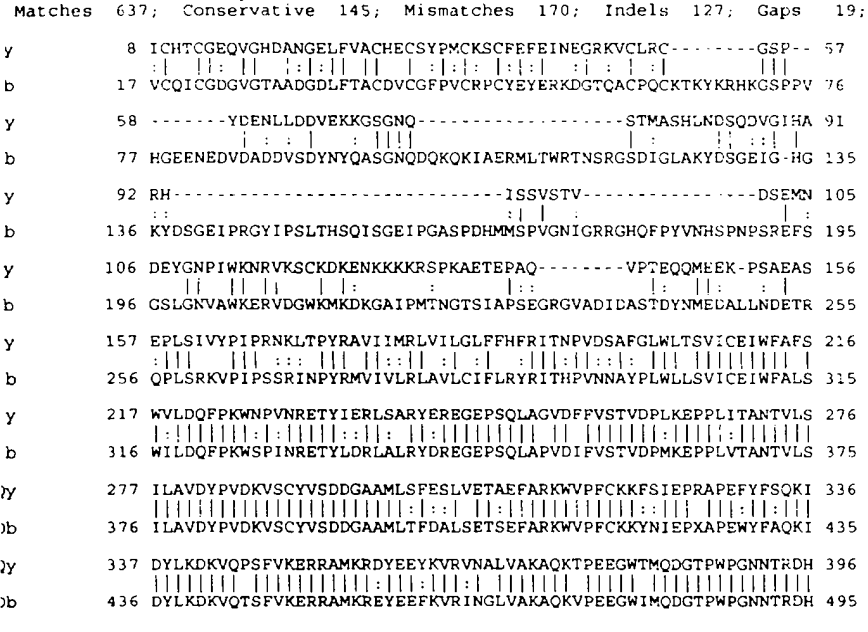
New genes which encode maize cellulose synthase polypeptides in plants useful for modulating the expression of cellulose synthase in plants and to produce transgenic plants expressing the novel protein -

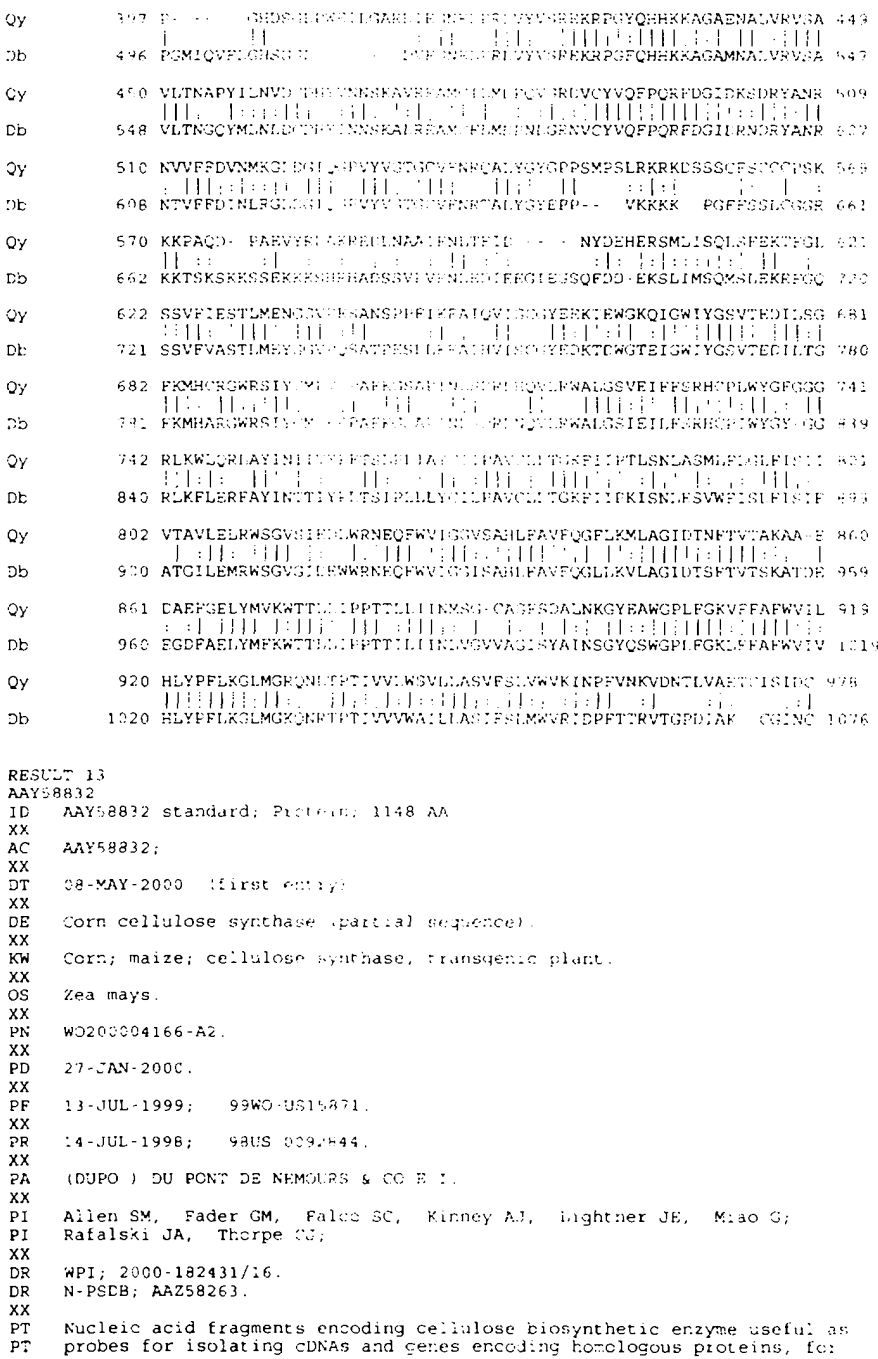
Claim 15; Page 196-199; 119pp; English.

The present sequence represents a maize cellulose synthase polypeptide. The cellulose synthase can be used for the improvement of stalk quality for improved stand or silage. It also provides an increased concentration of cellulose in the pericarp, hardening the kernel and improving its handling ability. The sequences are used to produce transgenic plants and seeds expressing the cellulose synthase. The polynucleotide is used for modulating, preferably increasing, the level of the synthase in a plant cell. The plants are preferably monocots. The polynucleotide is also used as a probe or primer in the detection quantitation or isolation of gene transcripts. The probes are useful in detecting deficiencies in the level of mRNA in screenings for desired transgenic plant, for detecting mutations in the gene, for monitoring upregulation of expression or changes in enzyme activity in screening assays of compounds, for detection of any number of allelic variants of the gene, or for use as molecular markers in plant breeding programs. The isolated nucleic acids of the present invention can also be used for recombinant expression of their encoded polypeptides or for use as immunogens in the preparation and/or screening of antibodies. The proteins can be employed in assays for enzyme agonists or antagonists of enzyme function or for use of immunogens or antigens to obtain antibodies specifically immunoreactive with a protein.

Sequence 1076 AA;

Query Match 64.1%; Score 3356.5; DB 21; Length 1076;
Best Local Similarity 59.0%; Pred. No. 1.4e-308;
Matches 637; Conservative 145; Mismatches 170; Indels 127; Gaps 19;





RESULT 13
AAY58832
ID AAY58832 standard; Protein; 1148 AA
XX
AC AAY58832;
XX
DT 08-MAY-2000 (first entry)
XX
DE Corn cellulose synthase (partial sequence).
XX
KW Corn; maize; cellulose synthase, transgenic plant.
XX
OS Zea mays.
XX
PN WO200004166-A2.
XX
PD 27-JAN-2000.
XX
PF 13-JUL-1999; 99WO-US15871.
XX
PR 14-JUL-1998; 98US-0092844.
XX
PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX
PI Allen SM, Fader GM, Falco SC, Kinney AJ, Lightner JE, Miao G;
PI Rafalski JA, Thorpe CJ;
XX
DR WPI; 2000-182431/16.
DR N-PSDB; AAZ58263.
XX
PT Nucleic acid fragments encoding cellulose biosynthetic enzyme useful as
PT probes for isolating cDNAs and genes encoding homologous proteins, for

```
PT   producing transgenic plants
XX
PS   Claim 12; Page 39-42; 93pp; English.
XX
CC   The present sequence is that of a portion of corn cellulose
CC   synthase (CS), as deduced from a contig of cDNA clones (see AAY58263).
CC   The invention relates to isolated CS nucleic acid fragments and to
CC   CS polypeptides.  It also relates to the construction of a chimeric
CC   gene encoding all or a portion of the CS, in sense or antisense
CC   orientation, where expression of the gene results in altered levels
CC   of the CS in transformed host cells.  The host cells can be used to
CC   screen compounds for their ability to inhibit CS activity.  CS
CC   nucleic acids are also useful for producing transgenic plants having
CC   altered levels of CS, and hence altered levels of fibre.  CS may
CC   also serve as a target for the development of novel herbicides.
XX
SQ   Sequence   1148 AA;

  Query Match             64.0%;  Score 3352;  DB 21,  Length 1148;
  Best Local Similarity   58.2%;  Pred. No. 4.1e-308;
  Matches  627;  Conservative  153;  Mismatches  178;  Indels  120;  Gaps   16;

Qy          1 MMESGAPICHTCGEQVGHDANGEIFVACHECSYPMCKSCFEFEINEGRKVCLRCGSPY-- 58
              :  :    :-  ||: :| : ::: ||| :|||:|:|: :|:|  || : | :| : :
Db         90 MDQRNGQVCQICGDDVGRNPDGEIFVACNECAFPICRDCYEYERREGTQNCPQCKTRFKR 149

Qy         59       -    --DENLLDDVEKKQSGNQSTMASHLNDSQEVGIHARHISSVSTVD------ 101
                        :|: :,|:| : :  :  : :' :|    :|: ':|     |
Db        150 LKGCARVPGDEEEDGVDDLENEFNWSDKHDSQYLAESM-- LHA-HMSYGRGADLDGVPQ 205

Qy        102 ------------SEMNDE------- -------          --------------- 107
                         :' |:
Db        206 PFEPIPNVPLLTNGQMVDDIPPEQHALVISEVSGGGKRIHRLPYADPNLPVQPRSMDPSK 265

Qy        108  -     YGNPIWKNRVKSCKDKENKKKKRSPKA---ETEPAQVPTEQQMEEKPSAEASEP 158
                 .|:  || |::| | :: :  :         : : |  :|                || :|
Db        266 DLAAYGYGSVAWKERMESWKQKQERMHQCRNDGGGDDGDDADLPLMD--------EARQP 317

Qy        159 LSIVYPIPRNKLTPYRAVIIMRLVILGLFFHFRITNPVESAFGLWLTSVICEIWFAFSWV 218
              '-   |:| ::: ||| :||:|''':|  ,'|:|: :||  :| ||| ||||||||' ||:
Db        318 LSRKIPLPSSQINPYRMIIIIRLVVLCFFLHYRVMHPVPDAFALWLISVICEIWFAMSWI 377

Qy        219 LDQFPKWNPVNRETYIERLSARYERDGEPSQLAGVDFFVSTVDPLKEPPLITANTVLSIL 278
              '|||||| .: ||||:'|| |.:.|' '  || '|||| ||||:|||:|||||||||
Db        378 LDQFPKWFPIERETYLDRLSLRFIKE'GEPSQLAPVDFFVSTVDPLKEPPLVTANTVLSIL 437

Qy        279 AVDYPVDKVSCYVSDDGAAMLTFEALSETSEFAKKWVPFCKKFSIEPRAPEFYFSQKIDY 338
              :| ||||||||'||||| '':|  :  ||:  |:|  ||| .|:|| ||||:|| |||||
Db        438 SVDYPVDKVSCYVSDDGAAMLTFEALSETSEFAKKWVPFCKKFSLEPRAPEWYFQQKIDY 497

Qy        339 LKDKVQPSFVKERRAMKRDYEEYKVRVNALVAKAQKTPEEGWTMQDGTPWPGNNTRDHPG 398
              ||| | |:||:||,||||:'|:| .:||:: ||| :||  '||||'|||||| ||:||
Db        498 LKDKVAPNFVRERRAMKREYEEFKVRINALVAKAQKVPEEGWTMQDGTPWPGNNVRDHPG 557

Qy        399 HDSGLPWEILGARDIEGNELPRLVYVSREKRPGYQHHKKAGAENALVRVSAVLTNAPYIL 458
                    :  '  |:|||||,,|||| :||'|' |; '||, '|||:,|||||||:|
Db        558 MIQVFLGQ-SGGHDVEGNELPRLVYVSREKRPGYNHHKKAGAMNALVRVSAVLTNAPYLL 616

Qy        459 NVDCDHYVNNSKAVREAMCILMDPQVGRDVCYVQFPQRFDGIDKSDRYANRNVVFFDVNM 518
              |:||.||:|||:|::|||| :||: :|: '|||' |||:'||: ||||||||||:||
Db        617 NLDCDHYINNSKAIKEAMCFMMDPLLGKKVCYVQFPQRFDGIDRHDRYANRNVVFFDINM 676

Qy        519 KGLDGIQGPVYVGTGCVFNRQALYGYGPPSMPSLRKRKDSS-- ---- -CFSCCC----- 566
              ||||:|||:|||||||| ||||| |       :  :        |  |||
Db        677 KGLDGIQGPIYVGTGCVFRRQALYGYDAPK  --TKKPPSRTCNCWPKWCICCCCFGNRK 732

Qy        567 PSKKKPAQDP--AEVYRDAKREDLNAAIFNITEILNY---DSIERSMLISQLSFEKTFGL 621
                ||      |   :: :  |:::  |  : ; |:'      :|:: :::|   || ||
Db        733 TKKKTKTSKPKFEKIKKLFKKKENQAIAYAIGEIDEAAPGAENEKASIVNQQKLEKKFGQ 792

Qy        622 SSVFIESTLMENGGVPESANSPFFIKEAIQVIRCGYEEKTEWGKQIGWIYGSVTEDILSG 681
              ||||: |||: ||  :|::    :|||   | |` |:|. || |:||||'|||: |:|
Db        793 SSVFVASTLLENGGTLKSASPASLLKEAIHVISCGYEDKTDWGKEIGWIYGSVTEDILTG 852

Qy        682 FKMHCRGWRSIYCMPVRPAFKGSAPINLSDRLHQVLRWALGSVEIFFSRHCPLWYGFGGG 741
              |||:| |||||:| | |||||||:|||||||||||||||:|||| |||||||:||
Db        853 FKMHCHGWRSIYCIPKRAAFKGSAPLNLSDRLHQVLRWALGSIEIFFSNHCPLWYGYGGG 912

Qy        742 RLKWLQRLAYINTIVYPFTSLPLIAYCTIPAVCLLTGKFIIPTLSNLASMLFLGLFISII 801
              ||:|:| :|||:||||:||:|:|:||||:||:|||||||| | |:|:||: |: ||| |
Db        913 -LKFLERFSYINSIVYPWTSIPLLAYCTLPAICLLTGKFITPELNNVASLWFMSLFICIF 971

Qy        802 VTAVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQGFLKMLAGIDTNFTVTAKAAED 861
              |::||:|||| |:| ||||||||||||:|||||||| ||::||:||:||||:|  :|
Db        972 ATSILEMRWSGVGIDDWWRNEQFWVIGGVSSHLFAVFQGLLKVIAGVDTSFTVTSKGGDD 1031

Qy        862 AEFGELYMVKWTTLLIPPTTLLIINMSG-CAGFSDALNKGYEAWGPLFGKVFFAFWVILH 920
              :| |||  |||||||||||||::|  |  || |:|:| |||:|||||||:|||||:|
Db       1032 EEFSELYTFKWTTLLIPPTTLLLLNFIGVVAGISNAINNGYESWGPLFGKLFFAFWVIVH 1091

Qy        921 LYPFLKGLMGRQNLTPTIVVLWSVLLASVFSLVWVKINPFVNKVDNTLVAETCISIDC 978
              ||||||||:|||| ||||:::||:||||:|||:||:|:||: | |  |: |    :||
Db       1092 LYPFLKGLVGRQNRTPTIVIVWSILLASIFSLLWVRIDPFLAKDDGPLLEE--CGLDC 1147


RESULT 14
AAG48733
ID   AAG48733 standard; Protein; 1065 AA.
XX
AC   AAG48733;
XX
DT   18-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 61573.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405 A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
```

20-MAY-1999; 99US-0135124.
21-MAY-1999; 99US-0135353.
24-MAY-1999; 99US-0135629.
25-MAY-1999; 99US-0136021.
27-MAY-1999; 99US-0136392.
28-MAY-1999; 99US-0136782.
01-JUN-1999; 99US-0137222.
03-JUN-1999; 99US-0137528.
04-JUN-1999; 99US-0137502.
07-JUN-1999; 99US-0137724.
08-JUN-1999; 99US-0138094.
10-JUN-1999; 99US-0138540.
10-JUN-1999; 99US-0138847.
14-JUN-1999; 99US-0139119.
16-JUN-1999; 99US-0139452.
16-JUN-1999; 99US-0139453.
17-JUN-1999; 99US-0139492.
18-JUN-1999; 99US-0139454.
18-JUN-1999; 99US-0139455.
18-JUN-1999; 99US-0139456.
18-JUN-1999; 99US-0139457.
18-JUN-1999; 99US-0139458.
18-JUN-1999; 99US-0139459.
18-JUN-1999; 99US-0139460.
18-JUN-1999; 99US-0139461.
18-JUN-1999; 99US-0139462.
18-JUN-1999; 99US-0139463.
18-JUN-1999; 99US-0139750.
18-JUN-1999; 99US-0139763.
21-JUN-1999; 99US-0139817.
22-JUN-1999; 99US-0139899.
23-JUN-1999; 99US-0140353.
23-JUN-1999; 99US-0140354.
24-JUN-1999; 99US-0140695.
28-JUN-1999; 99US-0140823.
29-JUN-1999; 99US 0140991.
30-JUN-1999; 99US-0141287.
01-JUL-1999; 99US-0141842.
01-JUL-1999; 99US-0142154.
02-JUL-1999; 99US-0142055.
06-JUL-1999; 99US-0142390.
08-JUL-1999; 99US-0142803.
09-JUL-1999; 99US-0142920.
12-JUL-1999; 99US-0142977.
13-JUL-1999; 99US-0143542.
14-JUL-1999; 99US-0143624.
15-JUL-1999; 99US-0144005.
16-JUL-1999; 99US-0144085.
16-JUL-1999; 99US-0144086.
19-JUL-1999; 99US-0144325.
19-JUL-1999; 99US-0144331.
19-JUL-1999; 99US-0144332.
19-JUL-1999; 99US-0144333.
19-JUL-1999; 99US-0144334.
19-JUL-1999; 99US-0144335.
20-JUL-1999; 99US-0144352.
20-JUL-1999; 99US-0144632.
20-JUL-1999; 99US-0144884.
21-JUL-1999; 99US-0144814.
21-JUL-1999; 99US-0145086.
21-JUL-1999; 99US-0145088.
22-JUL-1999; 99US-0145085.
22-JUL-1999; 99US-0145087.
22-JUL-1999; 99US-0145089.
22-JUL-1999; 99US-0145192.
23-JUL-1999; 99US-0145145.
23-JUL-1999; 99US-0145218.
23-JUL-1999; 99US-0145224.
26-JUL-1999; 99US-0145276.
27-JUL-1999; 99US-0145913.
27-JUL-1999; 99US-0145918.
27-JUL-1999; 99US-0145919.
28-JUL-1999; 99US-0145951.

PR 02-AUG-1999; 99US-0146486.
PR 02-AUG-1999; 99US-0146488.
PR 02-AUG-1999; 99US-0146489.
PR 03-AUG-1999; 99US-0146558.
PR 04-AUG-1999; 99US-0146824.
PR 04-AUG-1999; 99US-0146832.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147535.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148318.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148569.
PR 13-AUG-1999; 99US-0148884.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149479.
PR 18-AUG-1999; 99US-0149670.
PR 20-AUG-1999; 99US-0150[illegible].
PR 20-AUG-1999; 99US-0150[illegible].
PR 20-AUG-1999; 99US-0150[illegible].
PR 23-AUG-1999; 99US-0150[illegible].
PR 23-AUG-1999; 99US-0150[illegible].
PR 25-AUG-1999; 99US-0151[illegible].
PR 26-AUG-1999; 99US-0151884.
PR 27-AUG-1999; 99US-0151965.
PR 27-AUG-1999; 99US-0151966.
PR 27-AUG-1999; 99US-0151[illegible].
PR 30-AUG-1999; 99US-0152[illegible].
PR 31-AUG-1999; 99US-0152448.
PR 01-SEP-1999; 99US-0152830.
PR 07-SEP-1999; 99US-0153563.
PR 10-SEP-1999; 99US-0154070.
PR 13-SEP-1999; 99US-0154758.
PR 15-SEP-1999; 99US-0154918.
PR 16-SEP-1999; 99US-0154919.
PR 20-SEP-1999; 99US-0155775.
PR 22-SEP-1999; 99US-0156139.
PR 23-SEP-1999; 99US-0156486.
PR 24-SEP-1999; 99US-0156659.
PR 28-SEP-1999; 99US 0156498.
PR 29-SEP-1999; 99US-0157196.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158293.
PR 13-OCT-1999; 99US-0158294.
PR 13-OCT-1999; 99US 0158295.
PR 14-OCT-1999; 99US-0158329.
PR 14-OCT-1999; 99US-0158330.
PR 14-OCT-1999; 99US-0158331.
PR 14-OCT-1999; 99US-0158637.
PR 14-OCT-1999; 99US-0158638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT 1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.

PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 63.9%; Score 3350; DB 21; Length 1065;
Best Local Similarity 60.6%; Pred. No. 5.6e-308;
Matches 654; Conservative 123; Mismatches 185; Indels 118; Gaps 21;

Qy 2 MES- - GAPI--- --CHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEINEGRK 49
Db 1 MESEGETAGKPMKNIVPQTCQICSDNVGKTVDGDRFVACDICSFPVCRPCYEYERKDGNQ 60

Qy 50 VCLRC ------GSP ----YDENLLDDV- -------- ----------EKK 69
Db 61 SCPQCKTRYKRLKGSPAIPGDKDEDGLADEGTVEFNYPQKEKISERMLGWHLTRGKGEEM 120

Qy 70 GSGNQSTMASH-----LNDSQDV ----GIHARHISSVSTV- ------DSEMNDE---- 107
Db 121 GEPQYDKEVSHNHLPRLTSRQDTSGEFSAASPERLSVSSTIAGGKRLPYSSDVNQSPNRR 180

Qy 108 ------YGNPIWKNRVKSCKDKENKKK KRSPKARTEPAQVPTEQQM-----EEKPSAEA 155
Db 181 IVDPVGLGNVAWKERVDGWKMKQEKNTGPVSTQAASERGGVDIDASTDILADEALLNDEA 240

Qy 156 SEPLSIVYPIPRNKLTPYRAVIIMRLVILGLFFHFRITNPVDSAFGLWLTSVICEIWFAF 215
Db 241 RQPLSRKVSIPSSRINPYRMVIMLRLVILCLFLHYRITNPVHNAFALWLVSVICEIWFAL 300

Qy 216 SWVLDQFPKWNPVNRETYIERLSARYEREGEPSQLAGVDIFVSTVDPLKEPPLITANTVL 275
Db 301 SWILDQFPKWFPVNRETYLDRLAIRYDREGEPSQLAAVDIFVSTVDPLKEPPLVTANTVL 360

Qy 276 SILAVDYPVDKVSCYVSDDGAAMLSFESLVETAEFARKWVPFCKKFSIEPRAPEFYFSQK 335
Db 361 SILAVDYPVDKVSCYVSDDGAAMLSFESLAETSEFARKWVPFCKKYSIEPRAPEWYFAAK 420

Qy 336 IDYLKDKVQPSFVKERRAMKREYEEFKVRVNALVAKAQKTPEGWTMQDGTPWPGNNTRD 395
Db 421 IDYLKDKVQTSFVKDRRAMKREYEEFKIRINALVSKALKVPSEGWVMQDGTPWPGNNTRD 480

Qy 396 HP- - ---GHDSGLPWEILGARDIEGNELPRLVYVSREKRPGYQHHKKAGAENALVRVS 448
Db 481 HPGMIQVFLGQNGGL -- PAEGNELPRLVYVSREKRPGFQHHKKAGAMNALVRVS 532

Qy 449 AVLTNAPYILNVDCDHYVNNSKAVREAM
Db 533 AVLTNGPFILNLDCDHYIHNSKALREAM

Qy 509 RNVVFFDVNMKGLDGIQGPVYVGTGCVFNRQALYGYEPPRKV-- KHKKPSLLSKLCGGS 649
Db 593 RNTVFFDINLRGLDGIQGPVYVGTGCVFNRTALYGYEPPIKV-- KHKKPSLLSKLCGGS 649

Qy 569 KKK--PAQDPAEVYRDAKREDLNAAIFNLTEID---- NYDEHERSMLISQLSFEKTFG 620
Db 650 RKKNSKAKKESDKKKSGRHTDSTVPVFNLDDIEEGVEGAGFMD-EKALLMSQMSLEKRFG 708

Qy 621 LSSVFIESTLMENGGVPESANSPPFIKEAIQVIGCGYEEKTEWGKQIGWIYGSVTEDILS 680
Db 709 QSAVFVASTLMENGGVPPSATPENLLKEAIHVISCGYEDKSDWGMEIGWIYGSVTEDILT 768

Qy 681 GFKMHCRGWRSIYCMPVRPAFKGSAPINLSDRLHQVLRWALGSVEIFFSRHCPLWYGFGG 740
Db 769 GFKMHARGWRSIYCMPKLPAFKGSAPINLSDRLNQVLRWALGSVEILFSRHCPIWYGY-N 827

Qy 741 GRLKWLQRLAYINTIVYPFTSLPLIAYCTIPAVCLLTGKFIIPTLSNLASMLFLGLFISI 800
Db 828 GRLKFLERFAYVNTTIYPITSIPLLMYCTLPAVCLFTNQFIIPQISNIASIWFLSLFLSI 887

Qy 801 IVTAVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQGLKMLAGIDTNFTVTAKAA 859
Db 888 FATGILEMRWSGVGIDEWWRNEQFWVIGGVSAHLFAVFQGLKVLAGIDTNFTVTSKASD 947

Qy 860 EDAEFGELYMVKWTTLLIPPTTLLIINMSG-CAGFSDALNKGYEAWGPLFGKVFFAFWVI 918
Db 948 EDGDFAELYLFKWTTLLIPPTTLLIVNLVGVVAGVSYAINSGYQSWGPLFGKLFFAFWVI 1007

Qy 919 LHLYPFLKGLMGRQNLTPTIVVLWSVLLASVFSLVWVKINPFVNKVDNTLVAETCISIDC 978
Db 1008 VHLYPFLKGLMGRQNRTPTIVVVWSVLLASIFSLLWVRIDPFTSRVTGPDILE--CGINC 1065

RESULT 15
ABB93356
ID ABB93356 standard; Protein; 1065 AA.
XX
AC ABB93356;
XX
DT 31-MAY-2002 (first entry)
XX
DE Herbicidally active polypeptide SEQ ID NO 2567.
XX
KW Herbicidal; plant; agriculture; herbicide.
XX
OS Arabidopsis thaliana.
XX
PN WO200210210-A2.
XX
PD 07-FEB-2002.
XX
PF 28-AUG-2001; 2001WO-EP09892.
XX
PR 28-AUG-2001; 2001WO-EP09892.
XX
PA (FARB) BAYER AG.
XX
PI Tietjen K, Weidler M;
XX
DR WPI; 2002-269010/31.
XX
PT Identifying plant target proteins for herbicidally active compounds,
PT comprising aligning and comparing nucleic acid or amino acid sequences
PT from plant with nucleic acid or amino acid sequences from non-plant
PT organisms -
XX
PS Claim 5; SEQ ID NO 2567; 261pp + Sequence Listing; English.
XX
CC The invention relates to identifying target proteins
CC (ABB90790-ABB94016) for herbicidally active compounds, comprising
CC aligning and comparing nucleic acid or amino acid sequences from plant
CC with nucleic acid or amino acid sequences from non-plant organisms using
CC suitable search parameters, where plant sequences having an E-value
CC greater by a factor of 3 than the E-value of most similar non-plant
CC sequences are selected. The polypeptides or nucleic acids encoding them
CC are useful for identifying modulators. The identified modulators are
CC useful as herbicides.
XX
SQ Sequence 1065 AA;

Query Match 63.9%; Score 3350; DB 23; Length 1065;
Best Local Similarity 60.6%; Pred. No. 5.6e-308;
Matches 654; Conservative 123; Mismatches 185; Indels 118; Gaps 21;

Qy 2 MES-----GAPI-------CHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEINEGRK 49
Db 1 MESEGETAGKPMKNIVPQTCQICSDNVGKTVDGDRFVACDICSFPVCRPCYEYERKDGNQ 60

Qy 50 VCLRC--------GSP-----YDENLLDDV--------------------------EKK 69
Db 61 SCPQCKTRYKRLKGSPAIPGDKDEDGLADEGTVEFNYPQKEKISERMLGWHLTRGKGEEM 120

Qy 70 GSGNQSTMASH-----LNDSQDV-----GIHARHISSVSTV--------DSEMNDE---- 107
Db 121 GEPQYDKEVSHNHLPRLTSRQDTSGEFSAASPERLSVSSTIAGGKRLPYSSDVNQSPNRR 180

```
        108 ------YGNPIWKNRVKSCKDKENKKK-KRSPKAETEPAQVPTEQQM-----EEKPSAEA 155
        181 IVDPVGLGNVAWKERVDGWKMKQEKNTGPVSTQAASERGGVDIDASTDILADEALLNDEA 240

        156 SEPLSIVYPIPRNKLTPYRAVIIMRLVILGLFFHFRITNPVDSAFGLWLTSVICEIWFAF 215
        241 RQPLSRKVSIPSSRINPYRMVIMLRLVILCLFLHYRITNPVPNAFALWLVSVICEIWFAL 300

        216 SWVLDQFPKWNPVNRETYIERLSARYEREGEPSQLAGVDFFVSTVDPLKEPPLITANTVL 275
        301 SWILDQFPKWFPVNRETYLDRLALRYDREGEPSQLAAVDIFVSTVDPLKEPPLVTANTVL 360

        276 SILAVDYPVDKVSCYVSDDGAAMLSFESLVETAEFARKWVPFCKKFSIEPRAPEFYFSQK 335
        361 SILAVDYPVDKVSCYVSDDGAAMLSFESLAETSEFARKWVPFCKKYSIEPRAPEWYFAAK 420

        336 IDYLKDKVQPSFVKERRAMKRDYEEYKVRVNALVAKAQKTPEEGWTMQDGTPWPGNNTRD 395
        421 IDYLKDKVQTSFVKDRRAMKREYEEFKIRINALVSKALKCPEEGWVMQDGTPWPGNNTRD 480

        396 HP-------GHDSGLPWEILGARDIEGNELPRLVYVSREKRPGYQHHKKAGAENALVRVS 448
        481 HPGMIQVFLGQNGGL--------DAEGNELPRLVYVSREKRPGFQHHKKAGAMNALVRVS 532

        449 AVLTNAPYILNVDCDHYVNNSKAVREAMCILMDPQVGRDVCYVQFPQRFDGIDKSDRYAN 508
        533 AVLTNGPFILNLDCDHYINNSKALREAMCFLMDPNLGKQVCYVQFPQRFDGIDKNDRYAN 592

        509 RNVVFFDVNMKGLDGIQGPVYVGTGCVFNRQALYGYGPPSMPSLRKRKDSSSCFSCCCPS 568
        593 RNTVFFDINLRGLDGIQGPVYVGTGCVFNRTALYGYEPPIKV---KHKKPSLLSKLCGGS 649

        569 KKK--PAQDPAEVYRDAKREDLNAAIFNLTEID------NYDEHERSMLISQLSFEKTFG 620
        650 RKKNSKAKKESDKKKSGRHTDSTVPVFNLDDIEEGVEGAGFDD-EKALLMSQMSLEKRFG 708

        621 LSSVFIESTLMENGGVPESANSPPFIKEAIQVIGCGYEEKTEWGKQIGWIYGSVTEDILS 680
        709 QSAVFVASTLMENGGVPPSATPENLLKEAIHVISCGYEDKSDWGMEIGWIYGSVTEDILT 768

        681 GFKMHCRGWRSIYCMPVRPAFKGSAPINLSDRLHQVLRWALGSVEIFFSRHCPLWYGFGG 740
        769 GFKMHARGWRSIYCMPKLPAFKGSAPINLSDRLNQVLRWALGSVEILFSRHCPIWYGY-N 827

        741 GRLKWLQRLAYINTIVYPFTSLPLIAYCTIPAVCLLTGKFIIPTLSNLASMLFLGLFISI 800
        828 GRLKFLERFAYVNTTIYPITSIPLLMYCTLPAVCLFTNQFIIPQISNIASIWFLSLFLSI 887

        801 IVTAVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQGFLKMLAGIDTNFTVTAKAA- 859
        888 FATGILEMRWSGVGIDEWWRNEQFWVIGGVSAHLFAVFQGILKVLAGIDTNFTVTSKASD 947

        860 EDAEFGELYMVKWTTLLIPPTTLLIINMSG-CAGFSDALNKGYEAWGPLFGKVFFAFWVI 918
        948 EDGDFAELYLFKWTTLLIPPTTLLIVNLVGVVAGVSYAINSGYQSWGPLFGKLFFAFWVI 1007

        919 LHLYPFLKGLMGRQNLTPTIVVLWSVLLASVFSLVWVKINPFVNKVDNTLVAETCISIDC 978
       1008 VHLYPFLKGLMGRQNRTPTIVVVWSVLLASIFSLLWVRIDPFTSRVTGPDILE--CGINC 1065

earch completed: October 29, 2003, 18:29:00
ob time : 83.6132 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus p2n model

Run on: October 31, 2003, 23:20:30 ; Search time 5802.74 Seconds
(without alignments)
4096.305 Million cell updates/sec

Title: US-09-980-043-2
Perfect score: 978
Sequence: 1 MMESGAPICHTCGEQVGHDA.....PFVNKVLNTLVAETCISIDC 978

Scoring table: OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 12152248056 residues

Word size: 3

Total number of hits satisfying chosen parameters: 45562604

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post processing: Listing first 45 summaries

Command line parameters:
MODEL=frame+_p2n.model -DEV=xl% /cgn2 1/USPTO spool/US09980043/runat 29102003 16393 22677/app_query.fasta_1.1358 DB=EST QFMT=fastap -SUFFIX=oligo.rst -MINMATCH 0.1 -LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000 -USER=US09980043 aCGN_1_1_4125_srunat 29102003 163930 22677 -NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	145	14.8	484	12	BI128887	BI128887 G083P04Y
2	125	12.8	457	12	BI131891	BI131891 G127P18Y
3	110	11.2	571	13	BU895617	BU895617 X028D07 P
4	81	8.3	479	12	BI132326	BI132326 G135P19Y
5	78	8.0	731	13	BQ867361	BQ867361 QGD10J04.
6	70	7.2	287	12	BI129538	BI129538 G092P15Y
7	67	6.9	463	12	BI131187	BI131187 G116P69Y
8	57	5.8	614	29	BZ365529	BZ365529 ie20b12.g
c 9	57	5.8	770	29	BZ657620	BZ657620 OGCAX62TC
10	53	5.4	285	9	AI163338	AI163338 A040P09U
11	48	4.9	233	13	BU896785	BU896785 X045F11 P
12	48	4.9	565	9	AW691958	AW691958 NF051A02S
13	47	4.8	434	13	BU894130	BU894130 X003C01 P
14	47	4.8	476	13	BU897587	BU897587 X060B01 P
15	47	4.8	577	9	AW688489	AW688489 NF008C01S
16	47	4.8	735	13	BQ969585	BQ969585 QHB38P07.
17	44	4.5	435	12	BI127228	BI127228 G057P27Y
c 18	44	4.5	496	9	AI166604	AI166604 xylem.est
19	44	4.5	648	13	BU894292	BU894292 X005H01 P
20	44	4.5	683	13	BU897715	BU897715 X067B10 P
21	43	4.4	289	12	BI128439	BI128439 G076P02Y
c 22	43	4.4	368	9	AV560166	AV560166 AV560166
c 23	43	4.4	407	9	AV813548	AV813548 AV813548
c 24	43	4.4	411	9	AV817718	AV817718 AV817718
c 25	43	4.4	418	9	AV811500	AV811500 AV811500
c 26	43	4.4	421	9	AV816883	AV816883 AV816883
c 27	43	4.4	427	9	AV810887	AV810887 AV810887
c 28	43	4.4	606	14	CB819948	CB819948 EST 940 G
c 29	43	4.4	645	28	AQ967082	AQ967082 LERIP40TR
30	43	4.4	660	28	AQ967081	AQ967081 LERIP40TF
c 31	41	4.2	302	29	BX002729	BX002729 Arabidops
32	41	4.2	426	12	BI317423	BI317423 saf55d07.
33	41	4.2	573	12	BI784558	BI784558 saf56d07.
34	41	4.2	613	13	BU082548	BU082548 saq34c07.
35	40	4.1	335	14	D40691	D40691 RICS2805A R
36	40	4.1	398	10	BF586580	BF586580 FM1_29_A0
37	40	4.1	475	10	BF481793	BF481793 FM1_20_D0
38	40	4.1	494	10	BE355233	BE355233 DG1_113_C
39	40	4.1	554	12	BJ257238	BJ257238 BJ257238
40	40	4.1	556	12	BI643766	BI643766 IP1_56_G1
41	40	4.1	586	13	BQ578323	BQ578323 WHE0301_F
42	40	4.1	621	13	BQ506459	BQ506459 EST613874
43	40	4.1	658	10	BG644386	BG644386 EST506005
44	40	4.1	669	14	CD227355	CD227355 CCC1_50_D
45	40	4.1	686	13	BU005949	BU005949 QGG9712.y

ALIGNMENTS

RESULT 1
BI128887
LOCUS BI128887 484 bp mRNA linear EST 31-DEC-2001
DEFINITION G083P04Y Populus cambium cDNA library Populus tremula x Populus tremuloides cDNA, mRNA sequence.
ACCESSION BI128887
VERSION BI128887.1 GI:18012858
KEYWORDS EST.
SOURCE Populus tremula x Populus tremuloides
ORGANISM Populus tremula x Populus tremuloides
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids ; eurosids I; Malpighiales; Salicaceae; Populus.

```
FERENCE   1  (bases 1 to 484)
AUTHORS   Hertzberg,M., Aspeborg,H., Erlandsson,R., Bjorkbacka,H., Hiltonen
          ,T., Karlsson,J., Teeri,T., Gustafsson,P., Bahlerao,R., Jansson,S.,
          Nilsson,O., Sundberg,B., Nilsson,P., Uhlen,M., Sandberg,G. and
          Lundeberg,J.
TITLE     Gene expression in Populus
JOURNAL   Unpublished
MMENT     Contact: Erlandsson R
          Department of Biotechnology
          Royal Institute of Technology
          Teknikringen 30, Stockholm S-10044, Sweden
          Tel: 46 8 790 8287
          Fax: 46 8 245452
          Email: rikerl@biochem.kth.se.
ATURES              Location/Qualifiers
   source           1. .484
                    /organism="Populus tremula x Populus tremuloides"
                    /mol_type="mRNA"
                    /db_xref="taxon:47664"
                    /clone_lib="Populus cambium cDNA library"
                    /note="Organ: cambium"
SE COUNT      122 a    115 c    101 g    146 t
IGIN

ignment Scores:
ed. No.:              4.2e-131       Length:        484
ore:                  145.00         Matches:       145
rcent Similarity:     100.00%        Conservative:  0
st Local Similarity:  100.00%        Mismatches:    0
ery Match:            14.83%         Indels:        0
:                     12             Gaps:          0

-09-980-043-2 (1-978) x BI128887 (1-484)

/       742 ArgLeuLysTrpLeuGlnArgLeuAlaTyrIleAsnThrIleValTyrProPheThrSer 761
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
>         3 CGTCTTAAATGGCTCCAAAGGCTTGCGTATATAAACACCATTGTGTACCCATTTACATCC 62

/       762 LeuProLeuIleAlaTyrCysThrIleProAlaValCysLeuLeuThrGlyLysPheIle 781
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
>        63 CTCCCTCTCATTGCCTATTGCACAATTCCTGCAGTTTGTCTGCTCACCGGAAAATTCATC 122

/       782 IleProThrLeuSerAsnLeuAlaSerMetLeuPheLeuGlyLeuPheIleSerIleIle 801
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
>       123 ATACCAACGCTCTCAAACCTGGCAAGCATGCTGTTTCTTGGCCTCTTTATCTCCATCATT 182

y       802 ValThrAlaValLeuGluLeuArgTrpSerGlyValSerIleGluAspLeuTrpArgAsn 821
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
b       183 GTAACTGCGGTGCTTGAGCTAAGATGGAGCGGTGTCAGCATTGAAGATTTATGGCGTAAT 242

y       822 GluGlnPheTrpValIleGlyGlyValSerAlaHisLeuPheAlaValPheGlnGlyPhe 841
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
b       243 GAACAATTCTGGGTGATCGGAGGTGTTTCAGCCCATCTCTTTGCGGTCTTCCAGGGATTC 302

y       842 LeuLysMetLeuAlaGlyIleAspThrAsnPheThrValThrAlaLysAlaAlaGluAsp 861
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
b       303 TTAAAAATGTTGGCTGGCATCGATACGAACTTCACTGTCACAGCAAAAGCAGCCGAAGAT 362

y       862 AlaGluPheGlyGluLeuTyrMetValLysTrpThrThrLeuLeuIleProProThrThr 881
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
b       363 GCAGAATTTGGGGAGCTATATATGGTCAAGTGGACAACACTTTTGATTCCTCCAACAACA 422

y       882 LeuLeuIleIleAsn 886
            |||||||||||||||
b       423 CTTCTCATTATCAAT 437

ESULT 2
I131891
OCUS      BI131891               457 bp    mRNA    linear   EST 31-DEC-2001
EFINITION G127P18Y Populus cambium cDNA library Populus tremula x Populus
          tremuloides cDNA, mRNA sequence.
CCESSION  BI131891
ERSION    BI131891.1  GI:18015862
```

```
KEYWORDS    EST.
SOURCE      Populus tremula x Populus tremuloides
  ORGANISM  Populus tremula x Populus tremuloides
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids I; Malpighiales; Salicaceae; Populus.
REFERENCE   1  (bases 1 to 457)
  AUTHORS   Hertzberg,M., Aspeborg,H., Erlandsson,R., Bjorkbacka,H., Hiltonen
            ,T., Karlsson,J., Teeri,T., Gustafsson,P., Bahlerao,R., Jansson,S.,
            Nilsson,O., Sundberg,B., Nilsson,P., Uhlen,M., Sandberg,G. and
            Lundeberg,J.
  TITLE     Gene expression in Populus
  JOURNAL   Unpublished
COMMENT     Contact: Erlandsson R
            Department of Biotechnology
            Royal Institute of Technology
            Teknikringen 30, Stockholm S-10044, Sweden
            Tel: 46 8 790 8287
            Fax: 46 8 245452
            Email: rikerl@biochem.kth.se.
FEATURES              Location/Qualifiers
     source           1..457
                      /organism="Populus tremula x Populus tremuloides"
                      /mol_type="mRNA"
                      /db_xref="taxon:47664"
                      /clone_lib="Populus cambium cDNA library"
                      /note="Organ: cambium"
BASE COUNT      100 a    116 c    104 g    137 t
ORIGIN

Alignment Scores:
Pred. No.:              1.48e-111      Length:        457
Score:                  125.00         Matches:       151
Percent Similarity:     98.69%         Conservative:  0
Best Local Similarity:  98.69%         Mismatches:    0
Query Match:            12.29%         Indels:        2
DB:                     12             Gaps:          0

US-09-980-043-2 (1-978) x BI131891 (1-457)

Qy        688 GlyTrpArgSerIleTyrCysMetProValArgProAlaPheLysGlySerAlaProIle 707
              ||||||||||||.|||.|||.||||.|||.||||.|||.||||.||||.||||.||||.||||
Db          3 GGATGGAGATCAATCTACTGCATGCCCGTAAGGCCTGCGTTCAAAGGATCTGCACCCATC 62

Qy        708 AsnLeuSerAspArgProHisIleValLeuArgTrpAlaLeuGlySerValGluIlePhe 727
              ||||||||||||.|||.|||.||||.|||.||||.|||.||||.||||.||||.||||.||||
Db         63 AACCTGTCTGATAGACCTCACCAAGGTCCTCCGATGGGCTCTTGGTTCTGTGGAAATTTTC 122

Qy        728 PheSerArgHisCysProLeuTrpTyrGlyPheGlyGlyArgLeuLysTrpLeuGl 747
              ||||||||||||.|||.|||.||||.|||.||||.|||.||||.||||.||||.||||
Db        123 TTTAGCAGACACTGTCCCCTCTGGTATGGGTTTGGAGGAGGCCGTCTTAAATGGCTCCA 182

Qy        747 nArgLeuAlaTyrIleAsnThrIleValTyrProPheThrSerLeuProLeuIleAlaTy 767
              ||||.|||.|||.||||.|||.||||.|||.||||.|||.||||.||||.||||.||||.|
Db        183 AAGGCTTGCGTATATAAACACCATTGTGTACCCATTTACATCCCTCCCTCTCATTGCCTA 242

Qy        767 rCysThrIleProAlaValCysLeuLeuThrGlyLysPheIleIleProThrLeuSerAs 787
              |||.|||.|||.||||.|||.||||.|||.||||.|||.||||.||||.||||.||||.||
Db        243 TTGCACAATTCCTGCAGTTTGTCTGCTCACCGGAAAATTCATCATACCAACGCTCTCAAA 302

Qy        787 nLeuAlaSerMetLeuPheLeuGlyLeuPheIleSerIleIleValThrAlaValLeuGl 807
              ||||||.|||.|||.|||.|||.||||.|||.||||.|||.||||.||||.||||.||||
Db        303 CCTGGCAAGCATGCTGTTTCTTGGCCTCTTTATCTCCATCATTGTAACTGCGGTGCTTGA 362

Qy        807 uLeuArgTrpSerGlyValSerIleGluAspLeuTrpArg-AsnGluGlnPheTrpVal 827
              ||||||.|||.|||.|||.|||.||||.|||.||||.|||.|| ||||.||||.||||.||
Db        363 GCTAAGATGGAGCGGTGTCAGCATTGAAGATTTATGGCGTAAATGAACAATTCTGGGTGA 422

Qy        827 leGlyGlyValSerAlaHisLeuPheAlaValPhe 838
              |||||||.|||.|||.|||.|||.|||.|||
Db        423 TCGGAGGTGTATCAGCCCATCTCTTTGCGGTCTTC 457

RESULT 3
```

```
BU895617
LOCUS       BU895617                571 bp    mRNA    linear   EST 17-OCT-2002
DEFINITION  X028D07 Populus wood cDNA library Populus tremula x Populus
            tremuloides cDNA 5 prime, mRNA sequence.
ACCESSION   BU895617
VERSION     BU895617.1  GI:24106824
KEYWORDS    EST.
SOURCE      Populus tremula x Populus tremuloides
  ORGANISM  Populus tremula x Populus tremuloides
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids I; Malpighiales; Salicaceae; Populus.
REFERENCE   1  (bases 1 to 571)
  AUTHORS   Unneberg,P., Bhalerao,R.R., Jansson,S. and Sterky,F.
  TITLE     The poplar tree transcriptome: Analysis of expressed sequence tags
            from multiple libraries
  JOURNAL   Unpublished
COMMENT     Contact: BHALERAO RUPALI R.
            Umea Plant Science Center
            Department of Plant Physiology
            University of Umea, 901 87 Umea, Sweden
            Tel: +46 90 786 5279
            Fax: +46 90 786 6676
            Email: rupali.bhaleraoaplantphys.umu.se.
FEATURES             Location/Qualifiers
     source          1. .571
                     /organism="Populus tremula x Populus tremuloides"
                     /mol_type="mRNA"
                     /db_xref="taxon:47664"
                     /tissue_type="wood"
                     /clone_lib="Populus wood cDNA library"
BASE COUNT      145 a    131 c    122 g    143 t
ORIGIN

Alignment Scores:
Pred. No.:              8.25e-97        Length:         571
Score:                  110.00          Matches:        110
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            11.25%          Indels:         0
DB:                     13              Gaps:           0

US-09-980-043-2 (1-978) x BU895617 (1-571)

Qy        667 TyrGlyTrpIleTyrGlySerValThrGluAspIleLeuSer-GlyPheLysMetHisCys 686
              ||||:|||||| |.|||| .|||  ||| | | |  |||||||||||
Db        241 ACTGGTTTGGATCTATGGGTCAGTGACTTAGATAGCTTAAG-TGCTTCAAGATGCACTGC 300

Qy        687 ArgGlyTrpArgSerIleTyrCysMetPr ValArgPro-AlaPheLysGlySerAlaPro 706
              |:|||||||:|||||.|||| ||||  ||| ||  ||||||||||||||||||
Db        301 CGCGGATGGAGATCAATTTACTGCATGCCC TAAAGCCTG-TTCAAAGGATCTGCACCC 360

Qy        707 IleAsnLeuSerAspArgLeuHisGlnValLeuArgTrpAlaLeuGlySerValGluIle 726
              ||||||||||||||| ||||| ||||  ||||  |||||||||||||||||
Db        361 ATCAACCTGTCTGATAGATTGCACCAGGTCCTCCGATGGGCTCTTGGTTCTGTGGAAATT 420

Qy        727 PhePheSerArgHisCysProLeuTrpTyrGlyPheGlyGlyArgLeuLysTrpLeu 746
              |:|||||||||||||||| ||||| |||||||||||||||||||||||||||||||
Db        421 TTCTTTAGCAGACACTGTCCCCTCTGGTACGGGTTTGGAGGAGGCCGTCTTAAATGGCTC 480

Qy        747 GlnArgLeuAlaTyrIleAsnThrIleValTyrProPheThrSerLeuProLeuIleAla 766
              |||| ||||||||||||||||| |||||||||||||| |||||||||| ||||
Db        481 CAAAGGCTTGCGTATATAAACACCATTGTGTACCCATTTACATCCCTCCCTCTCATTGCC 540

Qy        767 TyrCysThrIleProAlaValCysLeuLeu 795
              ||||||| |||||||||||| ||||||||
Db        541 TATTGCACAATTCCTGCAGTTTGTCTGCTC 570

RESULT 4
BI132326
LOCUS       BI132326                479 bp    mRNA    linear   EST 31-DEC-2001
DEFINITION  G135P19Y Populus cambium cDNA library Populus tremula x Populus
            tremuloides cDNA, mRNA sequence.
ACCESSION   BI132326
VERSION     BI132326.1  GI:18016297
KEYWORDS    EST.
SOURCE      Populus tremula x Populus tremuloides
  ORGANISM  Populus tremula x Populus tremuloides
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids I; Malpighiales; Salicaceae; Populus.
REFERENCE   1  (bases 1 to 479)
  AUTHORS   Hertzberg,M., Aspeborg,H., Erlandsson,R., Bjorkbacka,H., Hiltonen
            ,T., Karlsson,J., Teeri,T., Gustafsson,P., Bahlerao,R., Jansson,S.,
            Nilsson,O., Sundberg,B., Nilsson,P., Uhlen,M., Sandberg,G. and
            Lundeberg,J.
  TITLE     Gene expression in Populus
  JOURNAL   Unpublished
COMMENT     Contact: Erlandsson R
            Department of Biotechnology
            Royal Institute of Technology
            Teknikringen 30, Stockholm S-10044, Sweden
            Tel: 46 8 790 8287
            Fax: 46 8 245452
            Email: rikerl@biochem.kth.se.
FEATURES             Location/Qualifiers
     source          1. .479
                     /organism="Populus tremula x Populus tremuloides"
                     /mol_type="mRNA"
                     /db_xref="taxon:47664"
                     /clone_lib="Populus cambium cDNA library"
                     /note="Organ: cambium"
BASE COUNT      111 a    116 c    113 g    139 t
ORIGIN

Alignment Scores:
Pred. No.:              1.69e-68        Length:         479
Score:                  81.00           Matches:        139
Percent Similarity:     98.58%          Conservative:   0
Best Local Similarity:  98.58%          Mismatches:     1
Query Match:            8.28%           Indels:         2
DB:                     12              Gaps:           0

US-09-980-043-2 (1-978) x BI132326 (1-479)

Qy        671 TyrGlySerValThrGluAspIleLeuSerGlyPheLysMetHisCysArgGlyTrpArg 690
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         23 TATGGGTCAGTCACTGAGGATATCTTAAGTGGCTTCAAGATGCACTGCCGCGGATGGAGA 82

Qy        691 SerIleTyrCysMetProValArgProAlaPheLysGlySerAlaProIleAsnLeuSer 710
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         83 TCAATTTACTGCATGCCCGTAAGGCCTGCGTTCAAAGGATCTGCACCCATCAACCTGTCT 142

Qy        711 AspArgLeuHisGlnValLeuArgTrpAlaLeuGly-SerValGluIlePhePheSerAr 730
              |||||||||||||||||||||||||||||||||||| |||||||||||||||||||   ||
Db        143 GATAGATTGCACCAGGTCCTCCGATGGGCTCTTGGTGTCTGTGGAAATTTTCTTTAC-AG 201

Qy        730 gHisCysProLeuTrpTyrGlyPheGlyGlyArgLeuLysTrpLeuGlnArgLeuAl 750
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        202 ACACTGTCCCCTCTGGTACGGGTTTGGAGGAGGCCGTCTTAAATGGCTCCAAAGGCTTGC 261

Qy        750 aTyrIleAsnThrIleValTyrProPheThrSerLeuProLeuIleAlaTyrCysThrIl 770
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        262 GTATATAAACACCATTGTGTACCCATTTACATCCCTCCCTCTCATTGCCTATTGCACAAT 321

Qy        770 eProAlaValCysLeuLeuThrGlyLysPheIleIleProThrLeuSerAsnLeuAlaSe 790
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        322 TCCTGCAGTTTGTCTGCTCACCGGAAAATTCATCATACCAACGCTCTCAAACCTGGCAAG 381

Qy        790 rMetLeuPheLeuGlyLeuPheIleSerIleIleValThrAlaValLeuGluLeuArgTr 810
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        382 CATGCTGTTTCTTGGCCTCTTTATCTCCATCATTGTAACTGCGGTGCTTGAGCTAAGATG 441

Qy        810 p 810
              :
Db        442 G 442
```

```
SULT 5
867361
CUS       BQ867361                731 bp    mRNA    linear   EST 14-AUG-2002
FINITION  QGD10J04.yg.ab1 QG_ABCDI lettuce salinas Lactuca sativa cDNA clone
          QGD10J04, mRNA sequence.
CESSION   BQ867361
RSION     BQ867361.1  GI:22253118
YWORDS    EST.
URCE      Lactuca sativa
 ORGANISM  Lactuca sativa
          Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
          Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
          asterids; campanulids; Asterales; Asteraceae; Cichorioideae;
          Cichorieae; Lactuca.
EFERENCE  1  (bases 1 to 731)
 AUTHORS  Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,
          Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegle,J., Ellison
          ,P., Kolkman,J., Slabaugh,M.S., Livingston,K., Zhou,Y., Lai,Z.,
          Church,S., Jackson,L. and Bradford,K.
 TITLE    Lettuce and Sunflower ESTs from the Compositae Genome Project
          http://compgenomics.ucdavis.edu/
 JOURNAL  Unpublished
OMMENT    Contact: Alexander Kozik (R.W.Michelmore)
          Department of Vegetable Crops, R.W.Michelmore Lab
          University of California at Davis (UCD)
          Asmundson Hall, UCD, Davis, CA 95616, USA
          Tel: 1-(530)-742-1742
          Fax: 1-(530)-752-9659
          Email: akozik@atgc.org (michelmore@vegmail.ucdavis.edu)
          belongs to contig QG_CA_Contig5167, see http://cgpdb.ucdavis.edu/
          for details.
          Plate: QGD10  row: J  column: 04.
EATURES             Location/Qualifiers
     source         1..731
                    /organism="Lactuca sativa"
                    /mol_type="mRNA"
                    /cultivar="Salinas"
                    /db_xref="taxon:4236"
                    /clone="QGD10J04"
                    /lab_host="E.coli"
                    /clone_lib="QG_ABCDI lettuce salinas"
                    /note="Vector: pBRcDNASfiAB; The library was constructed
                    from 10 different sources of RNA from a single genotype.
                    Separate cDNAs were generated using primers that
                    incorporated unique 5' and 3' tags to distinguish each
                    source of RNA. cDNAs were then pooled, size-fractionated,
                    directionally cloned into a custom medium-copy vector and
                    transformations made with four size classes to minimize
                    size bias. Details of each source of RNA and library
                    construction can be obtained at http://cgpdb.ucdavis.edu/
                    TAG_LIB=QG_ABCDI lettuce salinas
                    TAG_TISSUE=flowers pre-fertilized
                    TAG_SEQ=GCTTGACGGG"
ASE COUNT      209 a    149 c    160 g    213 t
RIGIN

lignment Scores:
red. No.:              2e-65          Length:         731
core:                  78.00          Matches:        145
ercent Similarity:     98.64%         Conservative:   0
est Local Similarity:  98.64%         Mismatches:     1
uery Match:            7.98%          Indels:         2
B:                     13             Gaps:           0

S-09-980-043-2 (1-978) x BQ867361 (1-731)

Qy      253 ValAspPhePheValSerThrValAspProLeuLysGluProProLeuIleThrAlaAsn 272
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      290 GTTGACTTCTTTGTGAGCACAGTTGACCCTTTGAAAGAGCCTCCTCTAATCACAGCCAAT 349

Qy      273 ThrValLeuSerIleLeuAlaValAspTyrProValAspLysValSerCysTyrValSer 292
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      350 ACTGTACTTTCAATGTTAGCTGTTGATTACCCGTTGATAAAGTCTCATGTTATGTCTGG 409

Qy      293 AspAspGlyAlaAlaMetLeuSerPheGluSerLeuValGluThrAlaGluPheAlaArg 312
            ||| |||.|||| ||| ||| ||| |||.|||||||||||||||||| |||||
Db      410 GATGATGGTGCTGCAATGCTTTCGTTTGAATCCCTTGTTGAAACCGCGGAGTTTGCAAGG 469

Qy      313 LysTrpValProLeuCysLysLysPheSerIleGluProArgAlaProGluPheTyrPh 332
            |||||||||||| ||| |||   ||.|||| ||||||||||||||||||||||||||
Db      470 AAGTGGGTCCCGTTTTGCAAACGTTTTCAATTGAGCCACGTGCACCTGAATTTTACTT 528

Qy      332 eSerGlnLysIleAsnTyrLeuLysAspLysValGlnProSerPheValLysGluArgAr 352
            ||||||||| ||| |||||||| |||||||||||||||||||||||||||||||||||
Db      529 CTCCCAAAAGATCAATTACTTGAAAGATAAAGTCCAACCTTCATTTGTGAAGGAACGTAG 588

Qy      352 gAlaMetLysArgAsnTyrGluGlnTyrLysValArgValAsnAlaLeuValAlaLysAl 372
            |||||||| ||| |||||| |||.|||| ||||||||||||||||||||||||||||||
Db      589 AGCAATGAAACGAAATTATGAGCAATACAAAGTGAGGGTAAATGCTTTAGTGGCTAAAGC 648

Qy      372 aGlnLysThrProSerGlyTrpLeuMetGlnAspGlyThrProTrpProGlyAsnAs 392
            ||||||||| |||| |||| |||| |||| |||||||||||||||||||||||||||
Db      649 TCAGAAAACTCCATCTGGATGGCTTATGCAAGATGGAACGCCCTGGCCTGGAAATAA 704

Qy      392 nThrArgAspPro                                           409
            ||| ||| |||                                            |||
Db      705 CACACGTGACCAA                                        TGA 727
```

```
RESULT 6
BI129538
LOCUS       BI129538                287 bp    mRNA    linear   EST 31 DEC 2001
DEFINITION  G292P15Y Populus tremula cDNA library Populus tremula x Populus
            tremuloides cDNA, mRNA sequence
ACCESSION   BI129538
VERSION     BI129538.1  GI:[illegible]
KEYWORDS    EST.
SOURCE      Populus tremula x Populus tremuloides
  ORGANISM  Populus tremula x Populus tremuloides
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids I; Malpighiales; Salicaceae; Populus.
REFERENCE   1  (bases 1 to 287)
  AUTHORS   Hertzberg,M., Aspeborg,H., Erlandsson,R., Bjorkbacka,H., Hiltonen
            ,T., Karlsson,J., Teeri,T., Gustafsson,P., Bahlerao,R., Jansson,S.,
            Nilsson,O., Sandberg,B., Nilsson,P., Uhlen,M., Sandberg,G. and
            Lundeberg,J.
  TITLE     Gene expression in Populus
  JOURNAL   Unpublished
COMMENT     Contact: Erlandsson R
            Department of Biotechnology
            Royal Institute of Technology
            Teknikringen 30, Stockholm S 10044, Sweden
            Tel: 46 8 790 8287
            Fax: 46 8 245452
            Email: rikerl@biochem.kth.se
FEATURES             Location/Qualifiers
     source          1..287
                     /organism="Populus tremula x Populus tremuloides"
                     /mol_type="mRNA"
                     /db_xref="taxon:47664"
                     /clone_lib="Populus cambium cDNA library"
                     /note="organ: cambium"
BASE COUNT       66 a     74 c     64 g     79 t
ORIGIN

Alignment Scores:
Pred. No.:              6.67e-58        Length:         287
Score:                  70.00           Matches:        70
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            7.16%           Indels:         0
DB:                     12              Gaps:           0

US-09-980-043-2 (1-978) x BI129538 (1-287)
```

```
Qy        698 ArgProAlaPheLysGlySerAlaProIleAsnLeuSerAspArgLeuHisGlnValLeu 717
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         65 AGGCCTGCATTCAAAGGATCTGCACCCATCAACCTGTCTGATAGATTGCACCAGGTCCTC 124

Qy        718 ArgTrpAlaLeuGlySerValGluIlePhePheSerArgHisCysProLeuTrpTyrGly 737
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        125 CGATGGGCTCTCGGTTCTGTGGAAATCTTCTTTAGCAGACACTGTCCCCTCTGGTACGGG 184

Qy        738 PheGlyGlyGlyArgLeuLysTrpLeuGlnArgLeuAlaTyrIleAsnThrIleValTyr 757
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        185 TTTGGAGGAGGCCGTCTTAAATGGCTCCAAAGGCTTGCGTATATAAACACCATTGTGTAC 244

Qy        758 ProPheThrSerLeuProLeuIleAlaTyr 767
              |||||||||||||||||||||||||||||
Db        245 CCATTTACATCCCTCCCTCTCATTGCCTAT 274

RESULT 7
BI131187
LOCUS       BI131187                 463 bp    mRNA    linear   EST 31-DEC 2001
DEFINITION  G116P69Y Populus cambium cDNA library Populus tremula x Populus
            tremuloides cDNA, mRNA sequence.
ACCESSION   BI131187
VERSION     BI131187.1  GI:18015158
KEYWORDS    EST.
SOURCE      Populus tremula x Populus tremuloides
  ORGANISM  Populus tremula x Populus tremuloides
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids I; Malpighiales; Salicaceae; Populus.
REFERENCE   1  (bases 1 to 463)
  AUTHORS   Hertzberg,M., Aspeborg,H., Erlandsson,R., Bjorkbacka,H., Hiltonen
            ,T., Karlsson,J., Teeri,T., Gustafsson,P., Bahlerao,R., Jansson,S.,
            Nilsson,O., Sundberg,B., Nilsson,P., Uhlen,M., Sandberg,G. and
            Lundeberg,J.
  TITLE     Gene expression in Populus
  JOURNAL   Unpublished
COMMENT     Contact: Erlandsson R
            Department of Biotechnology
            Royal Institute of Technology
            Teknikringen 30, Stockholm S-10044, Sweden
            Tel: 46 8 790 8287
            Fax: 46 8 245452
            Email: rikerl@biochem.kth.se
FEATURES             Location/Qualifiers
     source          1. .463
                     /organism="Populus tremula x Populus tremuloides"
                     /mol_type="mRNA"
                     /db_xref="taxon:[illegible]"
                     /clone_lib="Populus cambium cDNA library"
                     /note="Organ: cambium"
BASE COUNT      119 a    100 c    112 g    132 t
ORIGIN

Alignment Scores:
Pred. No.:              8.2e-55         Length:         463
Score:                  67.00           Matches:        [illegible]
Percent Similarity:     100.00%         Conservative:   [illegible]
Best Local Similarity:  100.00%         Mismatches:     [illegible]
Query Match:            6.85%           Indels:         [illegible]
DB:                     12              Gaps:           [illegible]

US-09-980-043-2 (1-978) x BI131187 (1-463)

Qy        667 IleGlyTrpIleTyrGlySerValThrGluAspIleLeuSerGlyPheLysMetHisCys 686
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        107 ATTGGTTGGATTTATGGGTCGGTCACTGAGGATATCTTAAGTGGATTCAAGATGCACTGC 166

Qy        687 ArgGlyTrpArgSerIleTyrCysMetProValArgProAlaPheLysGlySerAlaPro 706
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        167 CGAGGATGGAGATCAATATACTGCATGCCCGTAAGGCCAGCATTCAAAGGATCTGCACCC 226

Qy        707 IleAsnLeuSerAspArgLeuHisGlnValLeuArgTrpAlaLeuGlySerValGluIle 726
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        227 ATCAATCTGTCTGATAGATTGCACCAGGTTCTCCGGTGGGCTCTTGGATCTGTGGAAATT 286

Qy        727 PhePheSerArgHisCysPro 733
              |||||||||||||||||||||
Db        287 TTCTTTAGTAGACATTGTCCC 307

RESULT 8
BZ365529
LOCUS       BZ365529                 614 bp    DNA     linear   GSS 19-NOV-2002
DEFINITION  ie20b12.g2 WGS-ZmaysF (JM107 adapted methyl filtered) Zea mays
            genomic clone ie20b12 5', genomic survey sequence.
ACCESSION   BZ365529
VERSION     BZ365529.1  GI:25075022
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 614)
  AUTHORS   Rabinowicz,P.D., O'Shaughnessy,A.L., Balija,V., Dedhia,N.,
            Katzenburger,F., King,L., Miller,B., Muller,S., Nascimento,L.,
            Zutavern,T., McCombie,W.R. and Martienssen,R.A.
  TITLE     Genomic shotgun sequences from Zea mays (methyl-filtered)
  JOURNAL   Unpublished
COMMENT     Contact: W. Richard McCombie
            Lita Annenberg Hazen Genome Sequencing Center
            Cold Spring Harbor Laboratory
            PO Box 100, Cold Spring Harbor, NY 11724, USA
            Tel: 516 367 8884
            Fax: 516 367 8874
            Email: mccombie@cshl.org
            Plate: ie20  row: b  column: 12
            Seq primer: -21M13UnivRev
            Class: shotgun
            High quality sequence stop: 614.
FEATURES             Location/Qualifiers
     source          1. .614
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /cultivar="B73"
                     /db_xref="taxon:4577"
                     /clone="ie20b12"
                     /lab_host="JM107 or DH5a"
                     /clone_lib="WGS-ZmaysF (JM107 adapted methyl filtered)"
                     /note="Organ: immature ears; Site_1: Xba I; Site_2: Xba I;
                     The vector was digested with XbaI and one nucleotide was
                     added by fill in in the recessive 3' end. The genomic DNA
                     was nebulized, end repaired, adaptor ligated and size
                     fractionated using sephadex. The resulting fragments were
                     between 0.8 and 3 kb and were cloned into the vector
                     (.x/y reads in M13mp19, .b/g reads in pUC19). The same
                     ligation was transformed in either JM107 or DH5a. "
BASE COUNT      115 a    178 c    138 g    183 t
ORIGIN

Alignment Scores:
Pred. No.:              6.14e-45        Length:         614
Score:                  57.00           Matches:        57
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            5.83%           Indels:         0
DB:                     29              Gaps:           0

US-09-980-043-2 (1-978) x BZ365529 (1-614)

Qy        681 GlyPheLysMetHisCysArgGlyTrpArgSerIleTyrCysMetProValArgProAla 700
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        159 GGGTTTAAGATGCACTGCCGCGGCTGGAGATCCATCTACTGCATGCCGGTGAGGCCTGCA 218

Qy        701 PheLysGlySerAlaProIleAsnLeuSerAspArgLeuHisGlnValLeuArgTrpAla 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        219 TTCAAGGGATCAGCCCCGATCAACCTTTCCGACCGTCTTCACCAAGTTCTCCGGTGGGCT 278
```

```
        721 LeuGlySerValGluIlePhePheSerArgHisCysProLeuTrpTyrGly 737
            |||||||||||||||||||||||||||||||||||||||||||||||||||
        279 CTCGGTTCTGTCGAGATCTTCTTCAGTCGGCACTGCCCGCTGTGGTACGGA 329

SULT 9
657620/c
CUS       BZ657620               770 bp    DNA     linear   GSS 29-JAN-2003
FINITION  OGCAX62TC ZM_0.7_1.5_KB Zea mays genomic clone ZMMBMa0130L04,
          genomic survey sequence.
CESSION   BZ657620
RSION     BZ657620.1  GI:28130276
YWORDS    GSS.
URCE      Zea mays
RGANISM   Zea mays
          Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
          Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
          clade; Panicoideae; Andropogoneae; Zea.
FERENCE   1  (bases 1 to 770)
AUTHORS   Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick
          ,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T., Citek
          ,R.W., Nunberg,A., Robbins,D. and Lakey,N.
TITLE     Consortium for Maize Genomics
JOURNAL   Unpublished
MMENT     Other_GSSs: OGCAX62TM
          Contact: Cathy Whitelaw
          TIGR
          9712 Medical Center Drive, Rockville, MD 20850, USA
          Tel: 301-838-5843
          Fax: 301-838-0208
          Email: whitelaw@tigr.org
          Seq primer: TF
          Class: sheared ends.
ATURES              Location/Qualifiers
    source          1. .770
                    /organism="Zea mays"
                    /mol_type="genomic DNA"
                    /strain="B73"
                    /db_xref="taxon:4577"
                    /clone="ZMMBMa0130L04"
                    /clone_lib="ZM_0.7_1.5_KB"
                    /note="Vector: pBCSK-; Site_1: HincII; 0.7-1.5 kb
                    methylation filtered genomic DNA library"
ASE COUNT      216 a    184 c    211 g    159 t
RIGIN

ignment Scores:
red. No.:              7.27e-45       Length:         770
core:                  57.00          Matches:        57
ercent Similarity:     100.00%        Conservative:   0
est Local Similarity:  100.00%        Mismatches:     0
uery Match:            5.83%          Indels:         0
B:                     29             Gaps:           0

S-09-980-043-2 (1-978) x BZ657620 (1-770)

y       681 GlyPheLysMetHisCysArgGlyTrpArgSerIleTyrCysMetProValArgProAla 700
            !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
b       515 GGGTTTAAGATGCACTGCCGCGGCTGGAGATCCATCTACTGCATGCCGGTGAGGCCTGCA 456

y       701 PheLysGlySerAlaProIleAsnLeuSerAspArgLeuHisGlnValLeuArgTrpAla 720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
b       455 TTCAAGGGATCAGCCCCGATCAACCTTTCCGACCGTCTTCACCAAGTTCTCCGGTGGGCT 396

y       721 LeuGlySerValGluIlePhePheSerArgHisCysProLeuTrpTyrGly 737
            |||||||||||||||||||||||||||||||||||||||||||||||||||
b       395 CTCGGTTCTGTCGAGATCTTCTTCAGTCGGCACTGCCCGCTGTGGTACGGA 345

ESULT 10
I163338
OCUS      AI163338               285 bp    mRNA    linear   EST 03-DEC-1998
EFINITION A040P09U Hybrid aspen plasmid library Populus tremula x Populus
          tremuloides cDNA 5', mRNA sequence.

ACCESSION   AI163338
VERSION     AI163338.1  GI:3854623
KEYWORDS    EST.
SOURCE      Populus tremula x Populus tremuloides
  ORGANISM  Populus tremula x Populus tremuloides
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids I; Malpighiales; Salicaceae; Populus.
REFERENCE   1  (bases 1 to 285)
  AUTHORS   Sterky,F., Regan,S., Karlsson,J., Hertzberg,M., Rohde,A., Holmberg
            ,A., Amini,B., Bhalerao,R., Larsson,M., Villarroel,R., Van Montagu
            ,M., Sandberg,G., Olsson,O., Teeri,T.T., Boerjan,W., Gustafsson,P.,
            Uhlen,M., Sundberg,B. and Lundeberg,J.
  TITLE     Gene discovery in the wood forming tissues of poplar: Analysis of
            5,692 expressed sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 95 (22), 13330-13335 (1998)
  MEDLINE   99007314
   PUBMED   9789088
COMMENT     Contact: Sterky F
            Department of Biotechnology
            Royal Institute of Technology (KTH)
            Teknikringen 34, S-100 44 STOCKHOLM, Sweden
            Tel: +46 8 790 8[illegible]
            Fax: +46 8 24 54 [illegible]
            Email: fredrik[illegible]
            PCR PRimers
            FORWARD: AAAGGGPATGTGCTGCAAGGCG
            BACKWARD: GCTTCCPGCTCGTATGTTGTGTG
            Seq primer: CGTGTTAAAACGACGGCCAG
            High quality sequence stop: 285.
FEATURES            Location/Qualifiers
     source         1. .285
                    /organism="Populus tremula x Populus tremuloides"
                    /mol_type="mRNA"
                    /db_xref="taxon:47664"
                    /tissue_type="Cambial region"
                    /dev_stage="1.5 m actively growing tree"
                    /lab_host="E.coli"
                    /clone_lib="Hybrid aspen plasmid library"
                    /note="Vector: pBluescript SK; Site_1: SalI; Site_2: NotI;
                    Cambial region tissues, including developing xylem, the
                    meristematic  cambial zone and the developing and mature
                    phloem, was harvested  from 1.5 m actively growing trees
                    cDNA was prepared and cloned  into lambda gt22a  DNA was
                    isolated and subcloned into pBluescript SK using SalI and
                    NotI restriction enzymes."
BASE COUNT       75 a     56 c     74 g     79 t      1 others
ORIGIN

Alignment Scores:
Pred. No.:              2.83e-41       Length:         285
Score:                  53.00          Matches:        53
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            5.42%          Indels:         0
DB:                     9              Gaps:           0

US 09-980-043-2 (1-978) x AI163338 (1-285)

Qy      697 IleGlyTrpIleTyrGlySerValThrGluAspIleLeuSerGlyPheLysMetHisCys 686
            !!!!!!!!!! !!!! !!!!!!!!! !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
Db       62 ATTGGTTGGATTTAT GGTCGGTCACTGAGGATATCTTAAGTGGATTCAAGATGCACTGC 141

Qy      687 ArgGlyTrpArgSerIleTyrCysMetProValArgProAlaPheLysGlySerAlaPro 706
            !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
Db      142 CGAGGATGGAGATCAATATACTGCATGCCGGTAAGGCCAGCATTCAAAGGATCTGCACGC 201

Qy      707 IleAsnLeuSerAspArgLeuHisGlnValLeuArgTrp 719
            !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
Db      202 ATCAATCTGTCTGATAGATTGCACCAGGTTCTCCGGTGG 240

RESULT 11
BU896785
```

```
LOCUS       BU896785                 233 bp    mRNA    linear   EST 17-OCT-2002
DEFINITION  XC45F11 Populus wood cDNA library Populus tremula x Populus
            tremuloides cDNA 5 prime, mRNA sequence.
ACCESSION   BU896785
VERSION     BU896785.1  GI:24107992
KEYWORDS    EST.
SOURCE      Populus tremula x Populus tremuloides
  ORGANISM  Populus tremula x Populus tremuloides
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids I; Malpighiales; Salicaceae; Populus.
REFERENCE   1  (bases 1 to 233)
  AUTHORS   Unneberg,P., Bhalerao,R.R., Jansson,S. and Sterky,F.
  TITLE     The poplar tree transcriptome: Analysis of expressed sequence tags
            from multiple libraries
  JOURNAL   Unpublished
COMMENT     Contact: BHALERAO RUPALI R.
            Umea Plant Science Center
            Department of Plant Physiology
            University of Umea, 901 87 Umea, Sweden
            Tel: +46 90 786 5279
            Fax: +46 90 786 6676
            Email: rupali.bhaleraoesplantphys.umu.se.
FEATURES             Location/Qualifiers
     source          1. .233
                     /organism="Populus tremula x Populus tremuloides"
                     /mol_type="mRNA"
                     /db_xref="taxon:47664"
                     /tissue_type="wood"
                     /clone_lib="Populus wood cDNA library"
BASE COUNT       83 a     43 c     61 g     46 t
ORIGIN

Alignment Scores:
Pred. No.:                1.9e-36        Length:          233
Score:                    48.00          Matches:         48
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              4.91%          Indels:          0
DB:                       13             Gaps:            0

US-09-980-043-2 (1-978) x BU896785 (1-233)

Qy         125 AsnLysLysLysLysArgSerProLysAlaGluThrGluG  AlaGlnValProThrGlu 144
               ||||||||||||||||||||||||||||||||||||||||  |||||||||||||||||
Db          93 AACAAGAAGAAAAAAGAGAAGCCCAAGGCAGAAACCGAA  GCCCAGGTTCCTACGGAA 140

Qy         145 GlnGlnMetGluGlnLysGlnAlaLeuGlnGluPhe  ProSerIleValTyrPro 164
               |||||||||||||||||||||||||||||||||||||  |||||||||||||||||
Db         141 CAGCAGATGGAAGAGAAACAGGCTCTGCAGGAGTTC  CCTTCAATTGTTTATCCA 209

Qy         165 IleProArgAsnLysLeuThrPro 172
               ||||||||||||||||||||||||
Db         210 ATTCCACGCAACAAACTCACACCA 233

RESULT 12
AW691958
LOCUS       AW691958                 565 bp    mRNA    linear   EST 21-DEC-2000
DEFINITION  NF051A02ST1F1000 Developing stem Medicago truncatula cDNA clone
            NF051A02ST 5', mRNA sequence.
ACCESSION   AW691958
VERSION     AW691958.2  GI:11958349
KEYWORDS    EST.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 565)
  AUTHORS   He,X.-Z., Shadle,G., Scott,A.D., Harris,A.P., Gonzales,R.A., Bell
            ,C.J., Flores,H.R., Inman,J.T., Weiller,G.W., May,G.D. and Dixon
            ,R.A.
  TITLE     Expressed Sequence Tags from the Samuel Roberts Noble Foundation
            Medicago truncatula stem library
  JOURNAL   Unpublished
COMMENT     On Apr 14, 2000 this sequence version replaced gi:7566694.
            Contact: Dixon RA
            Plant Biology Division
            The Samuel Roberts Noble Foundation
            2510 Sam Noble Parkway, Ardmore, OK 73402, USA
            Tel: 580 221 7302
            Fax: 580 221 7380
            Email: radixon@noble.org
            Insert Length: 665   Std Error: 0.00
            Plate: 051  row: A  column: 02
            Seq primer: TCACACAGGAAACAGCTATGAC.
FEATURES             Location/Qualifiers
     source          1. .565
                     /organism="Medicago truncatula"
                     /mol_type="mRNA"
                     /db_xref="taxon:3880"
                     /clone="NF051A02ST"
                     /tissue_type="stem"
                     /dev_stage="Pooled developmental"
                     /clone_lib="Developing stem"
                     /note="Vector: Lambda Zap; Contains a mixture of
                     internodal stem segments"
BASE COUNT      171 a    105 c    140 g    149 t
ORIGIN

Alignment Scores:
Pred. No.:                3.68e-36       Length:          565
Score:                    48.00          Matches:         48
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              4.91%          Indels:          0
DB:                       9              Gaps:            0

US-09-980-043-2 (1-978) x AW691958 (1-565)

Qy         412 AspIleGluGlyAsnGluLeuProArgLeuValTyrValSerArgGluLysArgProGly 431
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         314 GACATAGAAGGAAACGAACTTCCAAGGTTGGTCTATGTTTCCAGAGAGAAAAGGCCAGGT 373

Qy         432 TyrGlnHisHisLysLysAlaGlyAlaGluAsnAlaLeuValArgValSerAlaValLeu 451
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         374 TACCAACACCACAAAAAGGCTGGTGCTGAAAATGCACTGGTGAGGGTGTCTGCAGTTCTC 433

Qy         452 ThrAsnAlaProTyrIleLeuAsn 459
               |||||||||||||||||||||||
Db         434 ACAAACGCTCCCTACATTCTCAAT 457

RESULT 13
BU894130
LOCUS       BU894130                 434 bp    mRNA    linear   EST 17-OCT-2002
DEFINITION  X003C01 Populus wood cDNA library Populus tremula x Populus
            tremuloides cDNA 5 prime, mRNA sequence.
ACCESSION   BU894130
VERSION     BU894130.1  GI:24105195
KEYWORDS    EST.
SOURCE      Populus tremula x Populus tremuloides
  ORGANISM  Populus tremula x Populus tremuloides
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids I; Malpighiales; Salicaceae; Populus.
REFERENCE   1  (bases 1 to 434)
  AUTHORS   Unneberg,P., Bhalerao,R.R., Jansson,S. and Sterky,F.
  TITLE     The poplar tree transcriptome: Analysis of expressed sequence tags
            from multiple libraries
  JOURNAL   Unpublished
COMMENT     Contact: BHALERAO RUPALI R.
            Umea Plant Science Center
            Department of Plant Physiology
            University of Umea, 901 87 Umea, Sweden
            Tel: +46 90 786 5279
```

```
          Fax: +46 90 786 6676
          Email: rupali.bhalerao@plantphys.umu.se.
ATURES              Location/Qualifiers
   source           1. .434
                    /organism="Populus tremula x Populus tremulcides"
                    /mol_type="mRNA"
                    /db_xref="taxon:47664"
                    /tissue_type="wood"
                    /clone_lib="Populus wood cDNA library"
SE COUNT      132 a     66 c    125 g    111 t
IGIN

ignment Scores:
ed. No.:              2.87e-35       Length:         434
ore:                  47.00          Matches:        47
rcent Similarity:     100.00%        Conservative:   0
st Local Similarity:  100.00%        Mismatches:     0
ery Match:            4.81%          Indels:         0
:                     13             Gaps:           0

-09-980-043-2 (1-978) x BU894130 (1-434)

          71 SerGlyAsnGlnSerThrMetAlaSerHisLeuAsnAspSerGluAspValGlyIleHis 90
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
         272 TCTGGCAATCAATCCACAATGGCATCTCACCTCAACGATTCTCAGGATGTCGGAATTCAT 331

          91 AlaArgHisIleSerSerValSerThrValAspSerGluMetAsnAspGluTyrGlyAsn 110
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
>        332 GCTAGACATATCAGTAGTGTGTCCACTGTGGATAGTGAAATGAATGATGAATATGGGAAT 391

/        111 ProIleTrpLysAsnArgVal 117
             |||||||||||||||||||||
>        392 CCAATTTGGAAGAATCGGGTG 412

ESULT 14
J897587
OCUS        BU897587               476 bp    mRNA    linear   EST 17-OCT 2002
EFINITION   XJ060B01 Populus wood cDNA library Populus tremula x Populus
            tremuloides cDNA 5 prime, mRNA sequence.
CCESSION    BU897587
ERSION      BU897587.1  GI:24108794
EYWORDS     EST.
OURCE       Populus tremula x Populus tremuloides
 ORGANISM   Populus tremula x Populus tremuloides
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids I; Malpighiales; Salicaceae; Populus.
EFERENCE    1  (bases 1 to 476)
 AUTHORS    Unneberg,P., Bhalerao,R.R., Jansson,S. and Sterky,F.
 TITLE      The poplar tree transcriptome: Analysis of expressed sequence tags
            from multiple libraries
 JOURNAL    Unpublished
OMMENT      Contact: BHALERAO RUPALI R.
            Umea Plant Science Center
            Department of Plant Physiology
            University of Umea, 901 87 Umea, Sweden
            Tel: +46 90 786 5279
            Fax: +46 90 786 6676
            Email: rupali.bhalerao@plantphys.umu.se.
EATURES             Location/Qualifiers
    source          1. .476
                    /organism="Populus tremula x Populus tremuloides"
                    /mol_type="mRNA"
                    /db_xref="taxon:47664"
                    /tissue_type="wood"
                    /clone_lib="Populus wood cDNA library"
ASE COUNT     114 a    108 c    103 g    151 t
ORIGIN

Alignment Scores:
Pred. No.:            3.08e-35       Length:         476
Score:                47.00          Matches:        87
Percent Similarity:   97.75%         Conservative:   0
```

```
Best Local Similarity:  [illegible]   Mismatches:   1
Query Match:            4.81%         Indels:       2
DB:                     13            Gaps:         0

US-09-980-043-2 (1-978) x [illegible] (1-476)

Qy        891 AlaGlyPheSerArgAlaLeuAspLysGlyTyrGluAlaTrpGlyProLeuPheGlyLys 910
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        183 GCTGGATTCTCTGGATCACTCAACAAAGGATATGAAGCATGGGGGCCTCTCTTTGGCAAG 242

Qy        911 ValPhePheAlaMetTrpValIleLeuHisLeuTyrProPheLeuLysGlyLeuMetGly 930
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        243 GTGTTCTTTGCTTT[illegible]GTGGATTCTCTCATCTTATCCATTCCTTAAAGGTCTAATGGGT 302

Qy        931 ArgGlnAsnLeu[illegible]ProThrIle[illegible]ValLeuTrpSerValLeuLeuAlaSerVal[illegible] 950
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        303 CGCCAAAA[illegible] 361

Qy        950 eSerLeuValTy[illegible]sIleAsn[illegible]eValAsnLysValAspAsnThrLeuVa.Al 970
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        362 CTCTCTTTCTTCT[illegible]AAGA[illegible]TAACAAAGTTGATAACACTTTGTTGC 421

Qy        971 aGluThrCys[illegible]cA[illegible] 990
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        422 GGAGACCTGCAT[illegible]ATCGATCT[illegible] 481

RESULT 15
AW688489
LOCUS       AW688489                  577 bp    mRNA    linear   EST 15-JUN-2000
DEFINITION  NF008C01ST1F10[illegible] developing stem Medicago truncatula cDNA clone
            NF008C01ST 5', mRNA sequence.
ACCESSION   AW688489
VERSION     AW688489.1  GI:[illegible]
KEYWORDS    EST.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to [illegible])
  AUTHORS   He,X.-Z., Shadle,G., Scott,A.D., Harris,A.R., Gonzales,E.A., Bell
            ,C.J., Flores,H.E., Inman,J.T., Weller,S.W., May,G.D. and Dixon
            ,R.A.
  TITLE     Expressed Sequence Tags from the Samuel Roberts Noble Foundation
            Medicago truncatula stem library
  JOURNAL   Unpublished
COMMENT     Contact: Dixon RA
            Plant Biology Division
            The Samuel Roberts Noble Foundation
            2510 Sam Noble Parkway, Ardmore, OK 73402, USA
            Tel: 580 221 7[illegible]
            Fax: 580 221 7[illegible]
            Email: radixons[illegible]e.org
            Insert Length: [illegible]  Std Error: [illegible]
            Plate: C08  row: [illegible]  column: [illegible]
            Seq primer: TCA[illegible]AGAAA[illegible]AGCTATGAC
FEATURES            Location/Qualifiers
     source         1. .577
                    /organism="Medicago truncatula"
                    /mol_type="mRNA"
                    /db_xref="taxon:3880"
                    /clone="NF008C01ST"
                    /tissue_type="stem"
                    /dev_stage="Pooled developmental"
                    /clone_lib="Developing stem"
                    /note="Vector: Lambda Zap; Contains a mixture of
                    internodal stem segments"
BASE COUNT      152 a    116 c    127 g    182 t
ORIGIN

Alignment Scores:
Pred. No.:            3.55e-35       Length:         577
```

Score:	47.00	Matches:	47
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	4.81%	Indels:	0
DB:	9	Gaps:	0

US-09-980-043-2 (1-978) x AW688489 (1-577)

```
Qy        251 ValAspPhePheValSerThrValAspProLeuLysGluProProLeuIleThrAlaAsn 272
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        383 GTTGATTTTTTTGTCAGTACAGTCGATCCTTTGAAAGAACCTACCACTGATTACAGCTAAT 442

Qy        273 ThrValLeuSerIleLeuAlaValAspTyrProValAspLysValSerCysTyrValSer 292
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        443 ACAGTGCTTTCTATCCTTGCGGTTGACTATCCAGTAGATAAAGTATCTTGTTATGTGTCA 502

Qy        293 AspAspGlyAlaAlaMetLeu 299
              |||||||||||||||||||||
Db        503 GATGATGGAGCAGCAATGCTC 523
```

Search completed: November 1, 2003, 04:44:50
Job time : 5813.74 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 29, 2003, 18:30:59 ; Search time 67.1547 Seconds
(without alignments)
5758.118 Million cell updates/sec

Title: US-09-980-043-2
Perfect score: 978
Sequence: 1 MMESGAPICHTCGEQVGHDA.....PFVNKVIGTLVAETCISIDC 978

Scoring table: CLIGO
Gapop 60.0 , Gapext 60.0

Searched: 830525 seqs, 258052604 residues

Word size : 0

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	978	100.0	978	10	O81368	O81368 populus tre
2	92	9.4	978	10	Q9AXK0	Q9axk0 zinnia eleg
3	61	6.2	974	10	P93155	P93155 gossypium h
4	61	6.2	974	10	Q8W1W0	Q8w1w0 gossypium h
5	47	4.8	504	10	Q9AXJ9	Q9axj9 zinnia eleg
6	47	4.8	505	10	Q9AXJ8	Q9axj8 zinnia eleg
7	47	4.8	958	10	Q9SN37	Q9sn37 arabidopsis
8	47	4.8	985	10	Q8LPK5	Q8lpk5 arabidopsis
9	47	4.8	985	10	Q9C5Z8	Q9c5z8 arabidopsis
10	44	4.5	56	10	Q9LLJ2	Q9llj2 populus tre
11	40	4.1	583	10	O48961	O48961 oryza sativ
12	40	4.1	1074	10	Q9LLI8	Q9lli8 zea mays (m
13	39	4.0	685	10	P93156	P93156 gossypium h
14	39	4.0	1043	10	Q9FNC3	Q9fnc3 arabidopsis
15	39	4.0	1055	10	Q8GZN8	Q8gzn8 arabidopsis
16	35	3.6	1065	10	Q9FHX6	Q9fhx6 arabidopsis
17	32	3.3	939	10	Q943H3	Q943h3 oryza sativ
18	32	3.3	1081	10	O65338	O65338 arabidopsis
19	32	3.3	1084	10	O48947	O48947 arabidopsis
20	32	3.3	1084	10	Q9FGF9	Q9fgf9 arabidopsis
21	30	3.1	1067	10	Q9XGX6	Q9xgx6 gossypium h
22	30	3.1	1076	10	Q9LLI5	Q9lli5 zea mays (m
23	30	3.1	1077	10	Q9LLI6	Q9lli6 zea mays (m
24	30	3.1	1079	10	Q9LLI1	Q9lli1 zea mays (m
25	29	3.0	346	10	Q8L778	Q8l778 arabidopsis
26	29	3.0	1069	10	Q9FIB9	Q9fib9 arabidopsis
27	29	3.0	1088	10	Q9SJ22	Q9sj22 arabidopsis
28	29	3.0	1091	10	Q93XQ1	Q93xq1 nicotiana a
29	28	2.9	507	10	Q93YP8	Q93yp8 arabidopsis
30	28	2.9	1063	10	Q9AV71	Q9av71 oryza sativ
31	26	2.7	1042	10	O81649	O81649 populus x c
32	25	2.6	366	10	Q94JQ6	Q94jq6 arabidopsis
33	24	2.5	821	10	Q9LLI7	Q9lli7 zea mays (m
34	24	2.5	1026	10	Q9XHP6	Q9xhp6 arabidopsis
35	24	2.5	1026	10	Q9SWW6	Q9sww6 arabidopsis
36	23	2.4	205	10	Q9SYX4	Q9syx4 gossypium h
37	23	2.4	205	10	Q9XFC6	Q9xfc6 gossypium r
38	23	2.4	205	10	Q9XFC5	Q9xfc5 gossypium r
39	23	2.4	205	10	Q9XFC4	Q9xfc4 gossypium h
40	23	2.4	205	10	Q9SYX5	Q9syx5 gossypium h
41	23	2.4	1075	10	Q9LLI9	Q9lli9 zea mays (m
42	23	2.4	1081	10	O48946	O48946 arabidopsis
43	22	2.2	1065	10	O48948	O48948 arabidopsis
44	22	2.2	1094	10	Q9LLI2	Q9lli2 zea mays (m
45	20	2.0	1032	10	Q8GSW2	Q8gsw2 populus tre

ALIGNMENTS

RESULT 1
O81368
ID O81368 PRELIMINARY; PRT; 978 AA.
AC O81368;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-JAN-1999 (TrEMBLrel. 09, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Secondary xylem cellulose synthase.
GN CELA.
OS Populus tremuloides (Quaking aspen).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Malpighiales; Salicaceae; Populus.
OX NCBI_TaxID=3693;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Secondary xylem;
RX MEDLINE=20344632; PubMed=10886769;
RA Wu L., Joshi C.P., Chiang V.L.;
RT "A xylem-specific cellulose synthase gene from aspen (Populus
RT tremuloides) is responsive to mechanical stress.";
RL Plant J. 22:495-502(2000).
DR EMBL; AF072131; AAD03417.1; -.
DR InterPro; IPR005150; Cellulose_synt.
DR Pfam; PF03552; Cellulose_synt; 1.
SQ SEQUENCE 978 AA; 110279 MW; C351B0E4975C6B8A CRC64;

Query Match 100.0%; Score 978; DB 10; Length 978;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 978; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MMESGAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEINEGRKVCLRCGSPYDE 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MMESGAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEINEGRKVCLRCGSPYDE 60

Qy         61 NLLDDVEKKGSGNQSTMASHLNDSQDVGIHARHISSVSTVDSEMNDEYGNPIWKNRVKSC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 NLLDDVEKKGSGNQSTMASHLNDSQDVGIHARHISSVSTVDSEMNDEYGNPIWKNRVKSC 120
```

```
 121 KDKENKKKKRSPKAETEPAQVPTEQQMEEKPSAEASEPLSIVYPIPRNKLTPYRAVIIMR 180
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 121 KDKENKKKKRSPKAETEPAQVPTEQQMEEKPSAEASEPLSIVYPIPRNKLTPYRAVIIMR 180

 181 LVILGLFFHFRITNPVDSAFGLWLTSVICEIWFAFSWVLDQFPKWNPVNRETYIERLSAR 240
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 181 LVILGLFFHFRITNPVDSAFGLWLTSVICEIWFAFSWVLDQFPKWNPVNRETYIERLSAR 240

 241 YEREGEPSQLAGVDFFVSTVDPLKEPPLITANTVLSILAVDYPVDKVSCYVSDDGAAMLS 300
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 241 YEREGEPSQLAGVDFFVSTVDPLKEPPLITANTVLSILAVDYPVDKVSCYVSDDGAAMLS 300

 301 FESLVETAEFARKWVPFCKKFSIEPRAPEFYFSQKIDYLKDKVQPSFVKERRAMKRDYEE 360
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 301 FESLVETAEFARKWVPFCKKFSIEPRAPEFYFSQKIDYLKDKVQPSFVKERRAMKRDYEE 360

 361 YKVRVNALVAKAQKTPEEGWTMQDGTPWPGNNTRDHPGHDSGLPWEILGARDIEGNELPR 420
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 361 YKVRVNALVAKAQKTPEEGWTMQDGTPWPGNNTRDHPGHDSGLPWEILGARDIEGNELPR 420

 421 LVYVSREKRPGYQHHKKAGAENALVRVSAVLTNAPYILNVDCDHYVNNSKAVREAMCILM 480
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 421 LVYVSREKRPGYQHHKKAGAENALVRVSAVLTNAPYILNVDCDHYVNNSKAVREAMCILM 480

 481 DPQVGRDVCYVQFPQRFDGIDKSDRYANRNVVFFDVNMKGLDGIQGPVYVGTGCVFNRQA 540
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 481 DPQVGRDVCYVQFPQRFDGIDKSDRYANRNVVFFDVNMKGLDGIQGPVYVGTGCVFNRQA 540

 541 LYGYGPPSMPSLRKRKDSSSCFSCCCPSKKKPAQDPAEVYRDAKREDLNAAIFNLTEIDN 600
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 541 LYGYGPPSMPSLRKRKDSSSCFSCCCPSKKKPAQDPAEVYRDAKREDLNAAIFNLTEIDN 600

 601 YDEHERSMLISQLSFEKTFGLSSVFIESTLMENGGVPESANSPPFIKEAIQVIGCCYEEK 660
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 601 YDEHERSMLISQLSFEKTFGLSSVFIESTLMENGGVPESANSPPFIKEAIQVIGCCYEEK 660

 661 TEWGKQIGWIYGSVTEDILSGFKMHCRGWRSIYCMPVRPAFKGSAPINLSDRLHQVLRWA 720
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 661 TEWGKQIGWIYGSVTEDILSGFKMHCRGWRSIYCMPVRPAFKGSAPINLSDRLHQVLRWA 720

 721 LGSVEIFFSRHCPLWYGFGGGRLKWLQRLAYINTIVYPFTSLPLIAYCTIPAVCLLTGKF 780
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 721 LGSVEIFFSRHCPLWYGFGGGRLKWLQRLAYINTIVYPFTSLPLIAYCTIPAVCLLTGKF 780

 781 IIPTLSNLASMLFLGLFISIIVTAVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQG 840
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 781 IIPTLSNLASMLFLGLFISIIVTAVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQG 840

 841 FLKMLAGIDTNFTVTAKAAEDAEFGELYMVKWTTLLIPPTTLLIINMSGCAGFSDALNKG 900
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 841 FLKMLAGIDTNFTVTAKAAEDAEFGELYMVKWTTLLIPPTTLLIINMSGCAGFSDALNKG 900

 901 YEAWGPLFGKVFFAFWVILHLYPFLKGLMGRQNLTPTIVVLWSVLLASVFSLVWVKINPF 960
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 901 YEAWGPLFGKVFFAFWVILHLYPFLKGLMGRQNLTPTIVVLWSVLLASVFSLVWVKINPF 960

 961 VNKVDNTLVAETCISIDC 978
     ||||||||||||||||||
 961 VNKVDNTLVAETCISIDC 978


ESULT 2
9AXK0
D   Q9AXK0      PRELIMINARY;      PRT;   978 AA.
C   Q9AXK0;
T   01-JUN-2001 (TrEMBLrel. 17, Created)
T   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
T   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
E   Cellulose synthase CesA-1.
S   Zinnia elegans.
C   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
C   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
C   Asteridae; campanulids; Asterales; Asteraceae; Asteroideae;
C   Heliantheae; Zinnia.
X   NCBI_TaxID=34249;
N   [1]
P   SEQUENCE FROM N.A.
X   MEDLINE=21438255; PubMed=11564447;
A   Haigler C.H., Ivanova-Datcheva M., Hogan P.S., Salnikov V.V.,
A   Hwang S., Martin K., Delmer D.P.;
T   "Carbon partitioning to cellulose synthesis.";
L   Plant Mol. Biol. 47:29-51(2001).
N   [2]
P   SEQUENCE FROM N.A.
A   Hwang S., Haigler C.H.;
L   Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
R   EMBL; AF323039; AAK11588.2; -.
R   InterPro; IPR005150; Cellulose_synt.
R   Pfam; PF03552; Cellulose_synt; 1.
Q   SEQUENCE   978 AA;  111 [illegible] MW;  [illegible] CRC64;

  Query Match          [illegible]; Score 1; DB 1; Length 978;
  Best Local Similarity  [illegible]; Pred. No. [illegible];
  Matches   98; Conservative  [illegible]; Mismatches  [illegible]; Indels  [illegible]; Gaps  [illegible];

Qy        801 FESLVETAEFARKW-VPFCKKFSIEPRAPEFYFSQKIDYLKDKVQPSFVKERRAMKRDYEE 860
              ||||.||| .|||   ||| .|||. |.|  |||.|.||||.|||| ..|||  |
Db        805 FESLVETAEFARKWSVPFKKFSIEPRAIEFYFSLKIDYLKDKVQPSFVKERRAMKRDYEE 864

Qy        861 YKVRVNALVAKAQKTPEEGWTMQDGTPWPGNN 892
              ||.|||.|||| |||| |||.|.|||.|
Db        865 YKVRVNALVAKAQKSPEESWTMQDGTPWPGNN 896


RESULT 3
P93155
ID   P93155      PRELIMINARY;      PRT;   974 AA.
AC   P93155;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Cellulose synthase.
GN   CELA1.
OS   Gossypium hirsutum (Upland cotton).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium.
OX   NCBI_TaxID=3635;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Acala SJ-2; TISSUE=Fiber;
RX   MEDLINE=97057296; PubMed=8901635;
RA   Pear J.R., Kawagoe Y., Schreckengost W.E., Delmer D.P., Stalker D.M.;
RT   "Higher plants contain homologs of the bacterial celA genes encoding
RT   the catalytic subunit of cellulose synthase.";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:12637-12642(1996).
DR   EMBL; U58283; AAB37766.1; -.
DR   InterPro; IPR005150; Cellulose_synt.
DR   Pfam; PF03552; Cellulose_synt; 1.
SQ   SEQUENCE   974 AA;  109[illegible] MW;  [illegible]CA2D[illegible]00A7 CRC64;

  Query Match          6.1%; Score 61; DB 10; Length 974;
  Best Local Similarity  100.0%; Pred. No. 7.2e-55;
  Matches   61; Conservative    0; Mismatches    0; Indels    0; Gaps    0;

Qy        799 SIIVTAVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQGFLKMLAGIDTNFTVTAKA 859
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        794 SIIVTAVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQGFLKMLAGIDTNFTVTAKA 853

Qy        859 A 859
              |
Db        854 A 854


RESULT 4
Q8W1W0
```

```
ID   Q8WIW0      PRELIMINARY;      PRT;   974 AA.
AC   Q8WIW0;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Cellulose synthase A4.
GN   CESA4.
OS   Gossypium hirsutum (Upland cotton).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium.
OX   NCBI_TaxID=3635;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21608363; PubMed=11743076;
RA   Kim H.J., Triplett B.A.;
RT   "Cotton Fiber Growth in Planta and in Vitro. Models for Plant Cell
RT   Elongation and Cell Wall Biogenesis.";
RL   Plant Physiol. 127:1361-1366(2001).
DR   EMBL; AF413210; AAL37710.1; -.
DR   InterPro; IPR005150; Cellulose_synt.
DR   Pfam; PF03552; Cellulose_synt; 1.
SQ   SEQUENCE   974 AA;  109499 MW;  F8AFB66ABA2B8071 CRC64;

  Query Match             6.2%;  Score 61;  DB 10;  Length 974;
  Best Local Similarity   100.0%;  Pred. No. 7.2e-55;
  Matches   61;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        799 SIIVTAVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQ[FLKMLAGIDTNFTVTAKA 858
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        794 SIIVTAVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQIFLKMLAGIDTNFTVTAKA 853

Qy        859 A 859
              |
Db        854 A 854


RESULT 5
Q9AXJ9
ID   Q9AXJ9      PRELIMINARY;      PRT;   504 AA.
AC   Q9AXJ9;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Cellulose synthase CesA 2 (Fragment).
OS   Zinnia elegans.
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC   Asteridae; campanulids; Asterales; Asteraceae; Asteroideae;
OC   Heliantheae; Zinnia.
OX   NCBI_TaxID=34245;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Haigler C.H., Ivanova-Datcheva M., Hogan P.S., Salnikov V.V.,
RA   Hwang S., Martin L.K., Delmer D.P.;
RT   "Carbon Partitioning to Cellulose Synthesis.";
RL   Plant Mol. Biol. 0:0-0(2001).
DR   EMBL; AF323040; AAK11589.1; -.
DR   InterPro; IPR005150; Cellulose_synt.
DR   Pfam; PF03552; Cellulose_synt; 1.
FT   NON_TER       1       1
SQ   SEQUENCE   504 AA;  56578 MW;  D9C3B5BE1565041D CRC64;

  Query Match             4.8%;  Score 47;  DB 10;  Length 504;
  Best Local Similarity   100.0%;  Pred. No. 2.6e-40;
  Matches   47;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        681 GFKMHCRGWRSIYCMPVRPAFKGSAPINLSDRLHQVLRWALGSVEIF 727
              |||||||||||||||||||||||||||||||||||||||||||||||
Db        207 GFKMHCRGWRSIYCMPVRPAFKGSAPINLSDRLHQVLRWALGSVEIF 253


RESULT 6
```

```
Q9AXJ8
ID   Q9AXJ8      PRELIMINARY;      PRT;   505 AA.
AC   Q9AXJ8;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Cellulose synthase CesA-3 (Fragment).
OS   Zinnia elegans.
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC   Asteridae; campanulids; Asterales; Asteraceae; Asteroideae;
OC   Heliantheae; Zinnia.
OX   NCBI_TaxID=34245;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Haigler C.H., Ivanova-Datcheva M., Hogan P.S., Salnikov V.V.,
RA   Hwang S., Martin L.K., Delmer D.P.;
RT   "Carbon Partitioning to Cellulose Synthesis.";
RL   Plant Mol. Biol. 0:0-0(2001).
DR   EMBL; AF323041; AAK11590.1; -.
DR   InterPro; IPR005150; Cellulose_synt.
DR   Pfam; PF03552; Cellulose_synt; 1.
FT   NON_TER       1       1
SQ   SEQUENCE   505 AA;  56478 MW;  9400BA486AD1B02D CRC64;

  Query Match             4.8%;  Score 47;  DB 10;  Length 505;
  Best Local Similarity   100.0%;  Pred. No. 2.7e-40;
  Matches   47;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        681 GFKMHCRGWRSIYCMPVRPAFKGSAPINLSDRLHQVLRWALGSVEIF 727
              |||||||||||||||||||||||||||||||||||||||||||||||
Db        207 GFKMHCRGWRSIYCMPVRPAFKGSAPINLSDRLHQVLRWALGSVEIF 253


RESULT 7
Q9SN37
ID   Q9SN37      PRELIMINARY;      PRT;   958 AA.
AC   Q9SN37;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Cellulose synthase-like protein.
GN   F28A21.190 OR AT4G18780.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Bevan M., Mueller M.W., Muendlein A., Felber R., Bancroft I.,
RA   Mewes H.W., Mayer K.F.X., Lemcke K., Schueller C.;
RL   Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   EU Arabidopsis sequencing project;
RL   Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Mueller M.W., Muendlein A., Felber R., Mewes H.W., Lemcke K.,
RA   Mayer K.F.X.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   EU Arabidopsis sequencing project;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL035526; CAB37463.1; -.
DR   EMBL; AL161549; CAB78880.1; -.
DR   InterPro; IPR005150; Cellulose_synt.
DR   Pfam; PF03552; Cellulose_synt; 1.
SQ   SEQUENCE   958 AA;  108520 MW;  B3836EF4BC697502 CRC64;

  Query Match             4.8%;  Score 47;  DB 10;  Length 958;
```

Best Local Similarity 100.0%; Pred. No. 4.7e-40;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

409 GARDIEGNELPRLVYVSREKRPGYQHHKKAGAENALVRVSAVLTNAP 455
408 GARDIEGNELPRLVYVSREKRPGYQHHKKAGAENALVRVSAVLTNAP 454

SULT 8
LPK5
Q8LPK5 PRELIMINARY; PRT; 985 AA
Q8LPK5;
01-OCT-2002 (TrEMBLrel. 22, Created)
01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
Cellulose synthase-like protein.
AT4G18780.
Arabidopsis thaliana (Mouse-ear cress).
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
eurosids II; Brassicales; Brassicaceae; Arabidopsis.
NCBI_TaxID=3702;
[1]
SEQUENCE FROM N.A.
Nguyen M., Karlin-Neumann G., Southwick A., Lam B., Miranda M.,
Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
Ecker J., Theologis A., Davis R.W.;
Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
EMBL; AY099636; AAM20487.1; -.
InterPro; IPR005150; Cellulose_synt.
Pfam; PF03552; Cellulose_synt; 1.
SEQUENCE 985 AA; 111521 MW; D299F357625768E0 CRC64;

Query Match 4.8%; Score 47; DB 10; Length 985;
Best Local Similarity 100.0%; Pred. No. 4.8e-40;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

409 GARDIEGNELPRLVYVSREKRPGYQHHKKAGAENALVRVSAVLTNAP 455
418 GARDIEGNELPRLVYVSREKRPGYQHHKKAGAENALVRVSAVLTNAP 464

ESULT 9
9C5Z8
D Q9C5Z8 PRELIMINARY; PRT; 985 AA.
C Q9C5Z8;
T 01-JUN-2001 (TrEMBLrel. 17, Created)
T 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
T 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
E Cellulose synthase catalytic subunit.
N IRX1.
S Arabidopsis thaliana (Mouse-ear cress).
C Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
C Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
C eurosids II; Brassicales; Brassicaceae; Arabidopsis.
X NCBI_TaxID=3702;
N [1]
P SEQUENCE FROM N.A.
X MEDLINE=21066939; PubMed=11148295;
A Taylor N.G., Laurie S., Turner S.R.;
T "Multiple Cellulose Synthase Catalytic Subunits Are Required for
T Cellulose Synthesis in Arabidopsis.";
L Plant Cell 12:2529-2540(2000).
R EMBL; AF267742; AAK08700.1; -.
R InterPro; IPR005150; Cellulose_synt.
R Pfam; PF03552; Cellulose_synt; 1.
Q SEQUENCE 985 AA; 111470 MW; 693C701B7B69AB25 CRC64;

Query Match 4.8%; Score 47; DB 10; Length 985;
Best Local Similarity 100.0%; Pred. No. 4.8e-40;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 409 GARDIEGNELPRLVYVSREKRPGYQHHKKAGAENALVRVSAVLTNAP 455
Db 418 GARDIEGNELPRLVYVSREKRPGYQHHKKAGAENALVRVSAVLTNAP 464

RESULT 10
Q9LLJ2
ID Q9LLJ2 PRELIMINARY; PRT; 56 AA
AC Q9LLJ2;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Cellulose synthase (Fragment).
GN CESA.
OS Populus tremuloides (Quaking aspen).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Malpighiales; Salicaceae; Populus.
OX NCBI_TaxID=3693;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Xylem;
RX MEDLINE=20344632; PubMed=10886269;
RA Wu L., Joshi C.P., Chiang V.L.;
RT "A xylem-specific cellulose synthase gene from aspen (Populus
RT tremuloides) is responsive to mechanical stress.";
RL Plant J. 22:495-502(2000).
DR EMBL; AF197911; AAF91184.1;
FT NON_TER 56 56
SQ SEQUENCE 56 AA; 6160 MW; 116FFFF0372E3402 CRC64;

Query Match 4.8%; Score 44; DB 10; Length 56;
Best Local Similarity 100.0%; Pred. No. 5.6e-38;
Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MMESGAPICHTCGEVGHDANGELFVACHECSYPMCKSCFEFEI 44
Db 1 MMESGAPICHTCGEVGHDANGELFVACHECSYPMCKSCFEFEI 44

RESULT 11
O48961
ID O48961 PRELIMINARY; PRT; 583 AA
AC O48961;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE RSW1-like cellulose synthase catalytic subunit (Fragment).
OS Oryza sativa (japonica cultivar-group)
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Nipponbare;
RX MEDLINE=98111412; PubMed=9445474;
RA Aricli T., Peng L., Betzner A.S., Burn J., Wittke W., Herth W.,
RA Camilleri C., Hofte H., Plazinski J., Birch R., Cork A., Glover J.,
RA Redmond J., Williamson R.E.;
RT "Molecular analysis of cellulose biosynthesis in Arabidopsis.";
RL Science 279:717-720(1998).
DR EMBL; AF030052; AAC39414.1; -.
DR Gramene; O48961; -.
DR InterPro; IPR005150; Cellulose_synt.
DR InterPro; IPR001841; Znf_ring.
DR Pfam; PF03552; Cellulose_synt; 1.
DR SMART; SM00184; RING; 1.
FT NON_TER 583 583
SQ SEQUENCE 583 AA; 65847 MW; 0E1E97BD73CAE16B CRC64;

```
  Query Match             4.1%;  Score 40;  DB 10;  Length 583;
  Best Local Similarity   100.0%;  Pred. No. 7.8e-33;
  Matches   40;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        256 FVSTVDPLKEPPLITANTVLSILAVDYPVDKVSCYVSDDG 295
              ||||||||||||||||||||||||||||||||||||||||
Db        354 FVSTVDPLKEPPLITANTVLSILAVDYPVDKVSCYVSDDG 393


RESULT 12
Q9LLI8
ID   Q9LLI8      PRELIMINARY;      PRT;  1074 AA.
AC   Q9LLI8;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Cellulose synthase-2.
GN   CESA-2
OS   Zea mays (Maize).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   PACCAD clade; Panicoideae; Andropogoneae; Zea.
OX   NCBI_TaxID=4577;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20398328; PubMed=10938450;
RA   Holland N., Holland D., Helentjaris T., Dhugga K.S.,
RA   Xoconostle-Cazares B., Delmer D.P.;
RT   "A comparative analysis of the plant cellulose synthase (CesA) gene
RT   family.";
RL   Plant Physiol. 123:1313-1324(2000)
DR   EMBL; AF200926; AAF89962.1; -
DR   InterPro; IPR005150; Cellulose_synt
DR   InterPro; IPR001841; Znf_ring
DR   Pfam; PF03552; Cellulose_synt; 1
DR   SMART; SM00184; RING; 1.
SQ   SEQUENCE   1074 AA;  121187 MW;  A116E88F0564E210 CRC64;

  Query Match             4.1%;  Score 40;  DB 10;  Length 1074;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-32;
  Matches   40;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        256 FVSTVDPLKEPPLITANTVLSILAVDYPVDKVSCYVSDDG 295
              ||||||||||||||||||||||||||||||||||||||||
Db        352 FVSTVDPLKEPPLITANTVLSILAVDYPVDKVSCYVSDDG 391


RESULT 13
P93156
ID   P93156      PRELIMINARY;      PRT;  685 AA.
AC   P93156;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Cellulose synthase (Fragment).
GN   CELA2.
OS   Gossypium hirsutum (Upland cotton).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium.
OX   NCBI_TaxID=3635;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Acala SJ-2; TISSUE=Fiber;
RX   MEDLINE=97057296; PubMed=8901635;
RA   Pear J.R., Kawagoe Y., Schreckengost W.E., Delmer D.P., Stalker D.M.;
RT   "Higher plants contain homologs of the bacterial celA genes encoding
RT   the catalytic subunit of cellulose synthase.";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:12637-12642(1996)
DR   EMBL; U58284; AAB37767.1; -
DR   InterPro; IPR005150; Cellulose_synt
DR   Pfam; PF03552; Cellulose_synt; 1
FT   NON_TER       1       1

SQ   SEQUENCE   685 AA;  77583 MW;  0B6A7037C0DA6959 CRC64;

  Query Match             4.0%;  Score 39;  DB 10;  Length 685;
  Best Local Similarity   100.0%;  Pred. No. 1e-31;
  Matches   39;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        417 ELPRLVYVSREKRPGYQHHKKAGAENALVRVSAVLTNAP 455
              |||||||||||||||||||||||||||||||||||||||
Db        105 ELPRLVYVSREKRPGYQHHKKAGAENALVRVSAVLTNAP 143


RESULT 14
Q9FNC3
ID   Q9FNC3      PRELIMINARY;      PRT;  1043 AA.
AC   Q9FNC3;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Cellulose synthase catalytic subunit-like protein.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Columbia;
RX   MEDLINE=98069011; PubMed=9405937;
RA   Kotani H., Nakamura Y., Sato S., Kaneko T., Asamizu E., Miyajima N.,
RA   Tabata S.;
RT   "Structural analysis of Arabidopsis thaliana chromosome 5. II.
RT   Sequence features of the regions of 1,044,062 bp covered by thirteen
RT   physically assigned P1 clones.";
RL   DNA Res. 4:291-300(1997).
DR   EMBL; AB006703; BAB09063.1; -.
DR   InterPro; IPR005150; Cellulose_synt.
DR   Pfam; PF03552; Cellulose_synt; 1.
SQ   SEQUENCE   1043 AA;  119127 MW;  B920242E2A12F295 CRC64;

  Query Match             4.0%;  Score 39;  DB 10;  Length 1043;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-31;
  Matches   39;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        698 RPAFKGSAPINLSDRLHQVLRWALGSVEIFFSRHCPLWY 736
              |||||||||||||||||||||||||||||||||||||||
Db        763 RPAFKGSAPINLSDRLHQVLRWALGSVEIFFSRHCPLWY 801


RESULT 15
Q8GZN8
ID   Q8GZN8      PRELIMINARY;      PRT;  1055 AA.
AC   Q8GZN8;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Cellulose synthase.
GN   IRX5.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Taylor N.G., Vickers K., Turner S.R.;
RT   "IRX5 encodes a cellulose synthase.";
RL   Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF458083; AAO15532.1; -.
SQ   SEQUENCE   1055 AA;  120108 MW;  9C2FC0DD289D4682 CRC64;

  Query Match             4.0%;  Score 39;  DB 10;  Length 1055;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-31;
  Matches   39;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
       698 RPAFKGSAPINLSDRLHQVLRWALGSVEIFFSRHCPLWY 736
           |||||||||||||||||||||||||||||||||||||||
       775 RPAFKGSAPINLSDRLHQVLRWALGSVEIFFSRHCPLWY 813

arch completed: October 29, 2003, 18:35:46
o time : 68.1547 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 29, 2003, 19:31:44 ; Search time 20.8085 Seconds
(without alignments)
2210.255 Million cell updates/sec

Title: US-09-980-043-2
Perfect score: 978
Sequence: 1 MMESGAPICHTCGEQVGHDA.....PFVNKVSNTLVAETCISIDC 978

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 127863 seqs, 47026705 residues

Word size : 0

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	8	0.8	1694	1	SN_MOUSE	Q62230 mus musculu
2	7	0.7	44	1	Y135_HELPY	O24948 helicobacte
3	7	0.7	65	1	YIME_BPPH1	P10437 bacteriopha
4	7	0.7	69	1	YVOC_VACCC	P20552 vaccinia vi
5	7	0.7	69	1	YVOC_VACCV	P04304 vaccinia vi
6	7	0.7	113	1	IMMC_ECOLI	P33637 escherichia
7	7	0.7	116	1	NUAM_PISSA	P24997 pisaster oc
8	7	0.7	117	1	YCX2_CHLMU	P75722 chlamydomon
9	7	0.7	145	1	DHI1_GOSHI	P09442 gossypium h
10	7	0.7	150	1	AVR1_CHICK	Q13153 gallus gall
11	7	0.7	150	1	AVR2_CHICK	P56732 gallus gall
12	7	0.7	150	1	AVR3_CHICK	P56733 gallus gall
13	7	0.7	150	1	AVR5_CHICK	P56735 gallus gall
14	7	0.7	150	1	AVR7_CHICK	P56736 gallus gall
15	7	0.7	155	1	DHB_CRAPL	P22239 craterostig
16	7	0.7	173	1	RT11_ACACA	P46758 acanthamoeb
17	7	0.7	181	1	Y4FF_RHISN	P55444 rhizobium s
18	7	0.7	184	1	KAD2_ANASP	Q8z0m3 anabaena sp
19	7	0.7	186	1	DH18_ARATH	P30185 arabidopsis
20	7	0.7	194	1	LPCA_BUCAP	Q5k9r9 buchnera ap
21	7	0.7	198	1	DEP_UREPA	Q9pq25 ureaplasma
22	7	0.7	217	1	TFOX_HAEIN	P43779 haemophilus
23	7	0.7	226	1	EXCY_RHIME	Q02731 rhizobium m
24	7	0.7	226	1	EXCY_RHISN	P14186 rhizobium s
25	7	0.7	257	1	VSP3_AGKAC	Q9i8x0 agkistrodon
26	7	0.7	260	1	VSP1_TRIFL	P05620 trimeresuru
27	7	0.7	261	1	SPRE_HUMAN	P35270 homo sapien
28	7	0.7	293	1	VP22_HSV7C	P52350 human herpe
29	7	0.7	312	1	YCHK_BACSU	P42243 bacillus su
30	7	0.7	317	1	FATC_VIBAN	P33737 vibrio angu
31	7	0.7	320	1	MCA2_CRIGR	Q9wvm7 cricetulus
32	7	0.7	320	1	MCA2_HUMAN	Q14155 homo sapien
33	7	0.7	331	1	YBHE_ECOLI	Q46926 escherichia
34	7	0.7	331	1	YBHE_ECOLI	P52697 escherichia
35	7	0.7	337	1	YBJS_ECOLI	P75821 escherichia
36	7	0.7	355	1	IL8A_RABIT	P21109 oryctolagus
37	7	0.7	360	1	IRKD_CAVPO	Q9qz65 cavia porce
38	7	0.7	360	1	IRKD_HUMAN	O60928 homo sapien
39	7	0.7	360	1	IRKD_RAT	O70617 rattus norv
40	7	0.7	360	1	JOSL_ARATH	O82391 arabidopsis
41	7	0.7	383	1	FLHB_BUCAI	P57334 buchnera ap
42	7	0.7	391	1	Y439_RICPR	Q9zda3 rickettsia
43	7	0.7	392	1	RECA_BIFBR	Q9s660 bifidobacte
44	7	0.7	394	1	CHSY_RAPSA	O22652 raphanus sa
45	7	0.7	395	1	CHS1_SINAL	P13416 sinapis alb

ALIGNMENTS

RESULT 1
SN_MOUSE
ID SN_MOUSE STANDARD; PRT; 1694 AA.
AC Q62230; O55216; Q62228; Q62229;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Sialoadhesin precursor (Sialic acid binding Ig-like lectin-1) (Siglec-
DE 1) (Sheep erythrocyte receptor) (SER).
GN SN OR SA.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=C57BL/6; TISSUE=Macrophage;
RX MEDLINE=95009950; PubMed=7925291;
RA Crocker P.R., Mucklow S., Bouckson V., McWilliam A., Willis A.C.,
RA Gordon S., Milon G., Kelm S., Bradfield P.;
RT "Sialoadhesin, a macrophage sialic acid binding receptor for
RT haemopoietic cells with 17 immunoglobulin-like domains.";
RL EMBO J. 13:4490-4503(1994).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=98051930; PubMed=9383289;
RA Mucklow S., Gordon S., Crocker P.R.;
RT "Characterization of the mouse sialoadhesin gene, Sn.";
RL Mamm. Genome 8:934-937(1997).
RN [3]
RP CHARACTERIZATION
RC STRAIN=C57BL/6; TISSUE=Spleen;
RX MEDLINE=91266893; PubMed=2050106;
RA Crocker P.R.;
RT "Purification and properties of sialoadhesin, a sialic acid-binding
RT receptor of murine tissue macrophages.";
RL EMBO J. 10:1661-1669(1991).
RN [4]
RP SIALIC ACID BINDING.
RX MEDLINE=95179521; PubMed=7533044;
RA Kelm S., Pelz A., Schauer R., Filbin M.T., Tang S., de Bellard M.E.,
RA Schnaar R.L., Mahoney J.A., Hartnell A., Bradfield P., Crocker P.R.;
RT "Sialoadhesin, myelin-associated glycoprotein and CD22 define a new
RT family of sialic acid-dependent adhesion molecules of the
RT immunoglobulin superfamily.";
RL Curr. Biol. 4:965-972(1994).
RN [5]
RP BINDING TO SPN.
RX MEDLINE=21136329; PubMed=11238599;
RA van den Berg T.K., Nath D., Ziltener H.J., Vestweber D., Fukuda M.,
RA van Die I., Crocker P.R.;
RT "CD43 functions as a T cell counterreceptor for the macrophage
RT adhesion receptor sialoadhesin (Siglec-1).";
RL J. Immunol. 166:3637-3640(2001).
RN [6]
RP X RAY CRYSTALLOGRAPHY (1.85 ANGSTROMS) OF 20-138.

```
MEDLINE=98325385; PubMed=9660955;
May A.P., Robinson R.C., Vinson M., Crocker P.R., Jones E.Y.;
"Crystal structure of the N-terminal domain of sialoadhesin in
complex with 3' sialyllactose at 1.85 A resolution.";
Mol. Cell 1:719-728(1998).
[7]
STRUCTURE OF 20-138 BY NMR, AND MUTAGENESIS OF TRP-21 AND ARG-116.
MEDLINE=99321481; PubMed=10393093;
Crocker P.R., Vinson M., Kelm S., Drickamer K.;
"Molecular analysis of sialoside binding to sialoadhesin by NMR and
site-directed mutagenesis.";
Biochem. J. 341:355-361(1999).
-!- FUNCTION: Macrophage-restricted adhesion molecule that mediates
    sialic-acid dependent binding to lymphocytes, including
    granulocytes, monocytes, natural killer cells, B-cells and CD8 T-
    cells (By similarity). Preferentially binds to alpha2,3-linked
    sialic acid. Binds to SPN/CD43 on T-cells. May play a role in
    hemopoiesis.
-!- SUBCELLULAR LOCATION: Type I membrane protein (isoform 1) and
    soluble (isoforms 2 and 3).
-!- ALTERNATIVE PRODUCTS:
    Event=Alternative splicing; Named isoforms=3;
    Name=1;
      IsoId=Q62230-1; Sequence=Displayed;
    Name=2;
      IsoId=Q62230-2; Sequence=VSP_002573, VSP_002574;
    Name=3;
      IsoId=Q62230-3; Sequence=VSP_002575, VSP_002576;
-!- TISSUE SPECIFICITY: Expressed by macrophages in various tissues.
    Highest expression in spleen and lymph node with lower amounts in
    lung, liver, bone marrow, heart and skin. No expression in thymus,
    kidney, brain or small intestine.
-!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. SIGLEC
    (SIALIC ACID BINDING IG-LIKE LECTIN) SUBFAMILY.
-!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
-!- SIMILARITY: Contains 16 immunoglobulin-like C2-type domains.
---------- ---------- ---------- ---
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
---------- ---------- ---
DR   EMBL; Z36293; CAA85290.1; -.
DR   EMBL; Z36233; CAA85268.1; -.
DR   EMBL; Z36234; CAA85269.1; -.
DR   EMBL; U92842; AAB95641.1; -.
DR   EMBL; U92833; AAB95641.1; JOINED.
DR   EMBL; U92834; AAB95641.1; JOINED.
DR   EMBL; U92836; AAB95641.1; JOINED.
DR   EMBL; U92837; AAB95641.1; JOINED.
DR   EMBL; U92838; AAB95641.1; JOINED.
DR   EMBL; U92839; AAB95641.1; JOINED.
DR   EMBL; U92840; AAB95641.1; JOINED.
DR   EMBL; U92841; AAB95641.1; JOINED.
DR   PIR; S50065; S50065.
DR   PDB; 1QFO; 16-APR-99.
DR   PDB; 1QFP; 16-APR-99.
DR   MGD; MGI:99668; Sn.
DR   GO; GO:0016021; C:integral to membrane; ISS.
DR   GO; GO:0005530; F:lectin; ISS.
DR   GO; GO:0016337; P:cell-cell adhesion; ISS.
DR   GO; GO:0007160; P:cell-matrix adhesion; ISS.
DR   GO; GO:0006954; P:inflammatory response; ISS.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003006; Ig_MHC.
DR   Pfam; PF00047; ig; 16.
DR   SMART; SM00408; IGc2; 5.
DR   PROSITE; PS50835; IG_LIKE; 14.
DR   PROSITE; PS00290; IG_MHC; 1.
KW   Cell adhesion; Lectin; Transmembrane; Signal; Glycoprotein;
KW   Immunoglobulin domain; Repeat; Alternative splicing; 3D-structure.
FT   SIGNAL       1     19
FT   CHAIN       20   1694       SIALOADHESIN.
FT   DOMAIN      20   1638       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM  1639   1659       POTENTIAL.
FT   DOMAIN    1660   1694       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      20    136       IG-LIKE V-TYPE.
FT   DOMAIN     143    235       IG-LIKE C2-TYPE 1.
FT   DOMAIN     239    321       IG-LIKE C2-TYPE 2.
FT   DOMAIN     326    406       IG-LIKE C2-TYPE 3.
FT   DOMAIN     416    503       IG-LIKE C2-TYPE 4.
FT   DOMAIN     509    594       IG-LIKE C2-TYPE 5.
FT   DOMAIN     602    701       IG-LIKE C2-TYPE 6.
FT   DOMAIN     704    781       IG-LIKE C2-TYPE 7.
FT   DOMAIN     796    890       IG-LIKE C2-TYPE 8.
FT   DOMAIN     894    972       IG-LIKE C2-TYPE 9.
FT   DOMAIN     978   1074       IG-LIKE C2-TYPE 10.
FT   DOMAIN    1082   1166       IG-LIKE C2-TYPE 11.
FT   DOMAIN    1171   1265       IG-LIKE C2-TYPE 12.
FT   DOMAIN    1344   1434       IG-LIKE C2-TYPE 13.
FT   DOMAIN    1441   1531       IG-LIKE C2-TYPE 14.
FT   DOMAIN    1441   1531       IG-LIKE C2-TYPE 15.
FT   DOMAIN    1533   1620       IG-LIKE C2-TYPE 16.
FT   SITE       827    829       CELL ATTACHMENT SITE (POTENTIAL).
FT   DISULFID    36    166       BY SIMILARITY.
FT   DISULFID    41     88
FT   DISULFID   160    218       BY SIMILARITY.
FT   DISULFID   263    306       BY SIMILARITY.
FT   DISULFID   347    391       BY SIMILARITY.
FT   DISULFID   434    492       BY SIMILARITY.
FT   DISULFID   532    576       BY SIMILARITY.
FT   DISULFID   625    685       BY SIMILARITY.
FT   DISULFID   725    770       BY SIMILARITY.
FT   DISULFID   813    872       BY SIMILARITY.
FT   DISULFID   921    969       BY SIMILARITY.
FT   DISULFID  1000   1062       BY SIMILARITY.
FT   DISULFID  1102   1144       BY SIMILARITY.
FT   DISULFID  1188   1236       BY SIMILARITY.
FT   DISULFID  1276   1319       BY SIMILARITY.
FT   DISULFID  1362   1421       BY SIMILARITY.
FT   DISULFID  1462   1508       BY SIMILARITY.
FT   DISULFID  1551   1610       BY SIMILARITY.
FT   CARBOHYD   159    159       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   266    266       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   299    299       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   340    340       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   509    509       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   583    584       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   693    693       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   722    722       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   737    737       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   882    882       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD  1089   1089       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD  1099   1099       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD  1246   1246       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD  1459   1459       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD  1473   1473       N-LINKED (GLCNAC...) (POTENTIAL).
FT   VARSPLIC   326    340       MAEVKMNTAGPVLEN...SESWMRLRGPVSGKH (in
FT                               isoform 2).
FT                               /FTId=VSP_002573.
FT   VARSPLIC   341   1694       Missing (in isoform 2).
FT                               /FTId=VSP_002574.
FT   VARSPLIC  1528   1599       YPPKTPTLIVEVEPQGGHQGILDCRVDSEPLAILTLHRGSQ
FT                               LVASNQLHDAPTKPEIRVTAPPNALRVEE...GEYRPIS
FT                               ALCLSLHLTGPYQAFSSAQSKGFIGKGLRTLASSLAGCMWF
FT                               VSMLGYFALKWRILLPFWDEYRR (in isoform 3).
FT                               /FTId=VSP_002575.
FT   VARSPLIC  1599   1694       Missing (in isoform 3).
FT                               /FTId=VSP_002576.
FT   MUTAGEN     21     21       W->Q: LOSS OF SIALIC ACID BINDING.
FT   MUTAGEN    116    116       R->A: LOSS OF SIALIC ACID BINDING.
FT   MUTAGEN    116    116       R->L: 10-FOLD LOSS IN AFFINITY TO SIALIC
```

```
  Query Match             0.8%;  Score 8;  DB 1;  Length 1694;
  Best Local Similarity   100.0%;  Pred. No. 17;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        945 LLASVFSL 952
              ||||||||
Db          9 LLASVFSL 16


RESULT 2
Y135_HELPY
ID   Y135_HELPY     STANDARD;      PRT;    44 AA.
AC   O24948;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Hypothetical protein HP0135/JHP0123 precursor.
GN   HP0135 OR JHP0123.
OS   Helicobacter pylori (Campylobacter pylori), and
OS   Helicobacter pylori J99 (Campylobacter pylori J99).
OC   Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC   Helicobacteraceae; Helicobacter.
OX   NCBI_TaxID=210, 85963;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=26695 / ATCC 700392;
RX   MEDLINE=97394467; PubMed=9252185;
RA   Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
RA   Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA   Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA   Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA   McKenney K., FitzGerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA   Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA   Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,
RA   Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA   Venter J.C.;
RT   "The complete genome sequence of the gastric pathogen Helicobacter
RT   pylori ";
RL   Nature 388:539-547(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=J99;
RX   MEDLINE=99120557; PubMed=9923682;
RA   Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA   Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA   Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA   Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA   Trust T.J.;
RT   "Genomic-sequence comparison of two unrelated isolates of the human
RT   gastric pathogen Helicobacter pylori ".
RL   Nature 397:176-180(1999).
CC   -!- ------  ---------------  ---  ---  ------- --------  ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -- ------- ---------------  --  ---  -- --------- ----
DR   EMBL; AE000535; AAD07212.1; -.
DR   EMBL; AE001451; AAD05700.1; -.
DR   PIR; G64536; G64536.
DR   TIGR; HP0135; -.
DR   PROSITE; PS00013; PROKAR_LIPOPROTEIN; UNKNOWN_1.
KW   Hypothetical protein; Signal; Complete proteome.
FT   SIGNAL        1     16       POTENTIAL.
FT   CHAIN        17     44       HYPOTHETICAL PROTEIN HP0135/JHP0123.
SQ   SEQUENCE   44 AA;  5023 MW;  8695773CB3807A7A CRC64;

  Query Match             0.7%;  Score 7;  DB 1;  Length 44;
  Best Local Similarity   100.0%;  Pred. No. 8.8;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         24 LFVACHE 30
              |||||||
Db         13 LFVACHE 19


RESULT 3
YIME_BPPH1
ID   YIME_BPPH1     STANDARD;      PRT;    65 AA.
AC   P10437;
DT   01-MAR-1989 (Rel. 10, Created)
DT   01-MAR-1989 (Rel. 10, Last sequence update)
DT   01-MAR-1989 (Rel. 10, Last annotation update)
DE   Hypothetical immunity region protein 14.
OS   Bacteriophage phi-105.
OC   Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;
OC   Lambda-like viruses.
OX   NCBI_TaxID=10717;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=86056972; PubMed=3934047;
RA   Cully D.F., Garro A.J.;
RT   "Nucleotide sequence of the immunity region of Bacillus subtilis
RT   bacteriophage phi 105: identification of the repressor gene and its
RT   mRNA and protein products.";
RL   Gene 38:153-164(1985).
CC   ---------------------------------------------------------------- ----------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------- ----------
DR   EMBL; M11920; AAA88395.1; -.
DR   PIR; D27234; IMBP14.
KW   Hypothetical protein.
SQ   SEQUENCE   65 AA;  7332 MW;  FA2B68EF47EA8AC6 CRC64;

  Query Match             0.7%;  Score 7;  DB 1;  Length 65;
  Best Local Similarity   100.0%;  Pred. No. 12;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        238 SARYERE 244
              |||||||
Db         23 SARYERE 29


RESULT 4
YVDC_VACCC
ID   YVDC_VACCC     STANDARD;      PRT;    69 AA.
AC   P20552;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical 7.3 kDa protein.
GN   D ORF C.
OS   Vaccinia virus (strain Copenhagen).
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC   Orthopoxvirus.
OX   NCBI_TaxID=10249;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91021027; PubMed=2219722;
RA   Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,
RA   Paoletti E.;
RT   "The complete DNA sequence of vaccinia virus.";
RL   Virology 179:247-266(1990).
RN   [2]
RP   COMPLETE GENOME.
RA   Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,
```

```
   Paoletti E.;
   "Appendix to 'The complete DNA sequence of vaccinia virus'.";
   Virology 179:517-563(1990).
   ---------------------------------------------- ------------------------
   This SWISS-PROT entry is copyright. It is produced through a collaboration
   between the Swiss Institute of Bioinformatics and the EMBL outstation -
   the European Bioinformatics Institute. There are no restrictions on its
   use by non-profit institutions as long as its content is in no way
   modified and this statement is not removed. Usage by and for commercial
   entities requires a license agreement (See http://www.isb-sib.ch/announce/
   or send an email to license@isb-sib.ch).
   ---------------------------------------------- ---- -------- ------- --
   EMBL; M35027; AAA48101.1; -.
   PIR; F42516; F42516.
   Hypothetical protein.
   SEQUENCE   69 AA;  7347 MW;  73DF495CBD7834BA CRC64;

Query Match             0.7%;  Score 7;  DB 1;  Length 69;
Best Local Similarity   100.0%;  Pred. No. 13;
Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

        760 TSLPLIA 766
            |||||||
         60 TSLPLIA 66


SULT 5
DC_VACCV
   YVDC_VACCV     STANDARD;      PRT;    69 AA.
   P04304;
   20-MAR-1987 (Rel. 04, Created)
   20-MAR-1987 (Rel. 04, Last sequence update)
   01-FEB-1991 (Rel. 17, Last annotation update)
   Hypothetical 7.3 kDa protein.
   D ORF C.
   Vaccinia virus (strain WR).
   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
   Orthopoxvirus.
   NCBI_TaxID=10254;
   [1]
   SEQUENCE FROM N.A.
   MEDLINE=86291159; PubMed=3739227;
   Niles E.G., Condit R.C., Caro P., Davidson K., Matusick L., Seto J.;
   "Nucleotide sequence and genetic map of the 16-kb vaccinia virus
   HindIII D fragment.";
   Virology 153:96-112(1986).
   ---------------------------------------------- ------------------------
   This SWISS-PROT entry is copyright. It is produced through a collaboration
   between the Swiss Institute of Bioinformatics and the EMBL outstation -
   the European Bioinformatics Institute. There are no restrictions on its
   use by non-profit institutions as long as its content is in no way
   modified and this statement is not removed. Usage by and for commercial
   entities requires a license agreement (See http://www.isb-sib.ch/announce/
   or send an email to license@isb-sib.ch).
   ---------------------------------------------- ------------------------
   EMBL; M15058; -; NOT_ANNOTATED_CDS.
   PIR; A03878; QQVZ7.
   Hypothetical protein.
   SEQUENCE   69 AA;  7321 MW;  B9BF494DBD701C18 CRC64;

 Query Match             0.7%;  Score 7;  DB 1;  Length 69;
 Best Local Similarity   100.0%;  Pred. No. 13;
 Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

        760 TSLPLIA 766
            |||||||
         60 TSLPLIA 66


ESULT 6
MM0_ECOLI
   IMM0_ECOLI     STANDARD;      PRT;    113 AA.
   P33637;
```

```
   01-FEB-1994 (Rel. 29, Created)
   01-FEB-1994 (Rel. 29, Last sequence update)
   16-OCT-2001 (Rel. 40, Last annotation update)
   Colicin E1 immunity protein (Imm) (Microcin E1 immunity protein).
   IMM.
   Escherichia coli.
   Plasmid IncFII R100.
   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
   Enterobacteriaceae; Escherichia.
   NCBI_TaxID=562;
   [1]
   SEQUENCE FROM N.A.
   MEDLINE=86319522; PubMed=3019092;
   Chatsubo H., Ryder T.B., Maeda Y., Armstrong K., Chatsubo E.;
   "DNA replication of the resistance plasmid R100 and its control.";
   Adv. Biophys. 21:115 (1986).
   -!- FUNCTION: THIS PROTEIN IS ABLE TO PROTECT A CELL, WHICH HARBORS
       THE PLASMID CoIE1 [MORE...]
   -!- [...]
   This SWISS-PROT entry is copyright. It is produced through a collaboration
   between the Swiss Institute of Bioinformatics and the EMBL outstation -
   the European Bioinformatics Institute. There are no restrictions on its
   use by non-profit institutions as long as its content is in no way
   modified and this statement is not removed. Usage by and for commercial
   entities requires a license agreement (See http://www.isb-sib.ch/announce/
   or send an email to license@isb-sib.ch).
   ---------------------------------------------- ------------------------
   EMBL; M26840; AAA26071.1;
   PIR; I64785; I64785.
   InterPro; IPR003081; Microcin
   Pfam; PF03526; Microcin; 1.
   PRINTS; PR01298; MICROCIN.
   KW   Bacteriocin immunity; Plasmid
   SQ   SEQUENCE   113 AA;  13000 MW;  4844710A0995A4C7 CRC64;

  Query Match             0.7%;  Score 7;  DB 1;  Length 113;
  Best Local Similarity   100.0%;  Pred. No. 20;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        794 LGLFISI 800
              |||||||
Db        104 LGLFISI 110


RESULT 7
NU3M_PISOC
ID   NU3M_PISOC     STANDARD;      PRT;    116 AA.
AC   P24997;
DT   01-MAR-1992 (Rel. 21, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   NADH-ubiquinone oxidoreductase chain 3 (EC 1.6.5.3).
GN   ND3.
OS   Pisaster ochraceus (Sea star).
OG   Mitochondrion.
OC   Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Asterozoa;
OC   Asteroidea; Forcipulatacea; Forcipulatida; Asteriidae; Pisaster.
OX   NCBI_TaxID=7612;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91012605; PubMed=1976816;
RA   Smith M.J., Hanfield D.F., Boteval K., Garski S., Kowbel D.J.;
RT   "Nucleotide sequence of nine protein coding genes and 22 tRNAs in the
RT   mitochondrial DNA of the sea star Pisaster ochraceus.";
RL   J. Mol. Evol. 31:135-204(1990).
CC   -!- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC   -!- SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 3 FAMILY.
CC   ---------------------------------------------- ---- ---- - -- -- ------ ------ - - - ---- ----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
```

```
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------- -- - --- ---- - - -- -------------- --- --
DR   EMBL; X55514; CAA39130.1; -.
DR   PIR; S14211; S14211.
DR   InterPro; IPR000440; Oxidored_q4.
DR   Pfam; PF00507; oxidored_q4; 1.
KW   Oxidoreductase; NAD; Ubiquinone; Mitochondrion.
SQ   SEQUENCE   116 AA;  13087 MW;  A8CAC5699C8185A CRC64;

  Query Match             0.7%;  Score 7;  DB 1;  Length 116;
  Best Local Similarity   100.0%;  Pred. No. 23;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        274 VLSILAV 280
              |||||||
Db          8 VLSILAV 14


RESULT 8
YCX2_CHLRE
ID   YCX2_CHLRE     STANDARD;      PRT;   117 AA.
AC   P05922;
DT   01-NOV-1988 (Rel. 09, Created)
DT   01-NOV-1988 (Rel. 09, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical 12.7 kDa protein in 16S rRNA region.
OS   Chlamydomonas reinhardtii.
OG   Chloroplast.
OC   Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC   Chlamydomonadaceae; Chlamydomonas.
OX   NCBI_TaxID=3055;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=86093659; PubMed=4080547;
RA   Schneider M., Darlix J. L., Erickson J., Rochaix J. D.;
RT   "Sequence organization of repetitive elements in the flanking regions
RT   of the chloroplast ribosomal unit of Chlamydomonas reinhardii.";
RL   Nucleic Acids Res. 13:8531-8541(1985).
CC   -!- MISCELLANEOUS: SIMILAR ORFS OF OPPOSITE POLARITY TO THE RRNA GENES
CC       HAVE ALSO BEEN FOUND IN HIGHER PLANTS IN THE RIBOSOMAL UPSTREAM
CC       REGION
CC   - - --- ---------------- --- ---------- ---------- ----------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------- -- - --- ---- - - -- -------------- --- --
DR   EMBL; X03269; CAA27023.1; -.
DR   PIR; S07933; S07933.
KW   Chloroplast; Hypothetical protein.
SQ   SEQUENCE   117 AA;  12752 MW;  4D0881C0794603B CRC64;

  Query Match             0.7%;  Score 7;  DB 1;  Length 117;
  Best Local Similarity   100.0%;  Pred. No. 23;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        266 PPLITAN 272
              |||||||
Db         48 PPLITAN 54


RESULT 9
DH11_GOSHI
ID   DH11_GOSHI     STANDARD;      PRT;   145 AA.
AC   P09442;
DT   01-MAR-1989 (Rel. 10, Created)
DT   01-MAR-1989 (Rel. 10, Last sequence update)
DT   01-DEC-1992 (Rel. 24, Last annotation update)
DE   Late embryogenesis abundant protein D-11 (LEA D-11).
OS   Gossypium hirsutum (Upland cotton).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium.
OX   NCBI_TaxID=3635;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Coker 201; TISSUE=Seed;
RA   Baker J., Steele C., Dure L. III;
RT   "Sequence and characterization of 6 Lea proteins and their genes from
RT   cotton.";
RL   Plant Mol. Biol. 11:277-291(1988).
CC   -!- FUNCTION: LEA PROTEIN ARE LATE EMBRYOGENESIS ABUNDANT IN HIGHER
CC       PLANT SEED EMBRYOS. THERE ARE TWO SUBSETS OF LEA PROTEINS (5A,
CC       AND 5B), THE FIRST ONES ARE EXPRESSED WHEN THE COTYLEDON WEIGHT
CC       REACH 80 MG AND THE SECOND SET ARE EXPRESSED ABOVE 100 MG. THE
CC       FUNCTION OF THOSE PROTEINS IS NOT KNOWN.
CC   -!- MISCELLANEOUS: THIS IS A SET 5B PROTEIN.
CC   -!- SIMILARITY: BELONGS TO THE PLANT DEHYDRIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M19379; AAA33073.1; -.
DR   EMBL; X13204; CAA31592.1; -.
DR   PIR; S04042; S04042.
DR   InterPro; IPR000167; Dehydrin.
DR   Pfam; PF00257; dehydrin; 1.
DR   PROSITE; PS00315; DEHYDRIN_1; 1.
DR   PROSITE; PS00823; DEHYDRIN_2; 1.
KW   Dehydrin; Seed; Seed embryo; Repeat.
FT   DOMAIN       61      69       POLY-SER.
SQ   SEQUENCE   145 AA;  15810 MW;  577791DDD7FDBC6A CRC64;

  Query Match             0.7%;  Score 7;  DB 1;  Length 145;
  Best Local Similarity   100.0%;  Pred. No. 25;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        106 DEYGNPI 112
              |||||||
Db         27 DEYGNPI 33


RESULT 10
AVR1_CHICK
ID   AVR1_CHICK     STANDARD;      PRT;   150 AA.
AC   O13153;
DT   15-JUL-1999 (Rel. 38, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Avidin-related protein 1 precursor.
GN   AVR1.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Oviduct;
RX   MEDLINE=88260103; PubMed=2838690;
RA   Keinaenen R.A., Laukkanen M.-L., Kulomaa M.S.;
RT   "Molecular cloning of three structurally related genes for chicken
RT   avidin.";
RL   J. Steroid Biochem. 30:17-21(1988).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=White leghorn; TISSUE=Oviduct;
```

```
MEDLINE=94170814; PubMed=8125122;
Keinaenen R.A., Wallen M.J., Kristo P.A., Laukkanen M.O.,
Toimela T.A., Helenius M.A., Kulomaa M.S.;
"Molecular cloning and nucleotide sequence of chicken avidin related
genes 1-5.";
Eur. J. Biochem. 220:615-621(1994).
-!- SIMILARITY: BELONGS TO THE AVIDIN/STREPTAVIDIN FAMILY.
---------------------------------------------------------------------
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
---------------------------------------------------------------------
EMBL; Z21611; -; NOT_ANNOTATED_CDS.
EMBL; Z97063; CAB09798.1; -.
PIR; S42201; S42201.
HSSP; P02701; 1RAV.
InterPro; IPR005469; Avidin.
InterPro; IPR005468; Avidin/str.
Pfam; PF01382; Avidin; 1.
PRINTS; PR00709; AVIDIN.
PROSITE; PS00577; AVIDIN; 1.
Biotin; Signal; Multigene family.
SIGNAL        1     24       POTENTIAL.
CHAIN        25    150       AVIDIN-RELATED PROTEIN 1.
BINDING      57     57       BIOTIN (NON-COVALENT) (BY SIMILARITY).
CARBOHYD     54     54       N-LINKED (GLCNAC. . .) (POTENTIAL).
CARBOHYD     67     67       N-LINKED (GLCNAC. . .) (POTENTIAL).
CARBOHYD     93     93       N-LINKED (GLCNAC. . .) (POTENTIAL).
SEQUENCE   150 AA;  16468 MW;  77548380B995B1F6 CRC64;

Query Match              0.7%;  Score 7;  DB 1;  Length 150;
Best Local Similarity  100.0%;  Pred. No. 25;
Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        744 KWLQRLA 750
              |||||||
Db        118 KWLQRLA 124


RESULT 11
AVR2_CHICK
ID   AVR2_CHICK     STANDARD;      PRT;   150 AA.
AC   P56732;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Avidin-related protein 2 precursor.
GN   AVR2.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=White leghorn; TISSUE=Oviduct;
RX   MEDLINE=94170814; PubMed=8125122;
RA   Keinaenen R.A., Wallen M.J., Kristo P.A., Laukkanen M.O.,
RA   Toimela T.A., Helenius M.A., Kulomaa M.S.;
RT   "Molecular cloning and nucleotide sequence of chicken avidin-related
RT   genes 1-5.";
RL   Eur. J. Biochem. 220:615-621(1994).
CC   -!- SIMILARITY: BELONGS TO THE AVIDIN/STREPTAVIDIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z21554; -; NOT_ANNOTATED_CDS.
DR   EMBL; Z21535; -; NOT_ANNOTATED_CDS.
DR   PIR; S42202; S42202.
DR   HSSP; P02701; 1RAV.
DR   InterPro; IPR005469; Avidin.
DR   InterPro; IPR005468; Avidin/str.
DR   Pfam; PF01382; Avidin; 1.
DR   PRINTS; PR00709; AVIDIN.
DR   PROSITE; PS00577; AVIDIN; FALSE_NEG.
KW   Biotin; Signal; Multigene family.
FT   SIGNAL        1     24       POTENTIAL.
FT   CHAIN        25    150       AVIDIN-RELATED PROTEIN 2.
FT   BINDING      57     57       BIOTIN (NON-COVALENT) (BY SIMILARITY).
FT   CARBOHYD     67     67       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     93     93       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   150 AA;  16464 MW;  76705E2B1A93A66D CRC64;

  Query Match              0.7%;  Score 7;  DB 1;  Length 150;
  Best Local Similarity  100.0%;  Pred. No. 25;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        744 KWLQRLA 750
              |||||||
Db        118 KWLQRLA 124


RESULT 12
AVR3_CHICK
ID   AVR3_CHICK     STANDARD;      PRT;   150 AA.
AC   P56733;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Avidin-related protein 3 precursor.
GN   AVR3.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=White leghorn; TISSUE=Oviduct;
RX   MEDLINE=94170814; PubMed=8125122;
RA   Keinaenen R.A., Wallen M.J., Kristo P.A., Laukkanen M.O.,
RA   Toimela T.A., Helenius M.A., Kulomaa M.S.;
RT   "Molecular cloning and nucleotide sequence of chicken avidin-related
RT   genes 1-5.";
RL   Eur. J. Biochem. 220:615-621(1994).
CC   -!- SIMILARITY: BELONGS TO THE AVIDIN/STREPTAVIDIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z21612; -; NOT_ANNOTATED_CDS.
DR   EMBL; Z21536; -; NOT_ANNOTATED_CDS.
DR   PIR; S42203; S42203.
DR   HSSP; P02701; 1RAV.
DR   InterPro; IPR005468; Avidin/str.
DR   Pfam; PF01382; Avidin; 1.
DR   PROSITE; PS00577; AVIDIN; 1.
KW   Biotin; Signal; Multigene family.
FT   SIGNAL        1     24       POTENTIAL.
FT   CHAIN        25    150       AVIDIN-RELATED PROTEIN 3.
FT   BINDING      57     57       BIOTIN (NON-COVALENT) (BY SIMILARITY).
```

```
FT   CARBOHYD      93      93       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     141     141       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   150 AA;  16657 MW;  84DC03926A68B21C CRC64;

  Query Match             0.7%;  Score 7;  DB 1;  Length 150;
  Best Local Similarity   100.0%;  Pred. No. 25;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        744 KWLQRLA 750
              |||||||
Db        118 KWLQRLA 124


RESULT 13
AVR6_CHICK
ID   AVR6_CHICK     STANDARD;      PRT;   150 AA.
AC   P56735;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Avidin-related protein 6 precursor.
GN   AVR6.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Rhode Island;
RX   MEDLINE=21070478; PubMed=11167523;
RA   Ahlroth M.K., Kola E.H., Ewald D., Masabanda J., Sazanov A., Fries R.,
RA   Kulomaa M.S.;
RT   "Characterization and chromosomal localization of the chicken avidin
RT   gene family.";
RL   Anim. Genet. 31:367-375(2000).
CC   -!- SIMILARITY: BELONGS TO THE AVIDIN/STREPTAVIDIN FAMILY.
CC   ------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to licensewisb-sib.ch).
CC   ------------------------------------------------------------------------
DR   EMBL; AJ237658; CAB39893.1; -.
DR   HSSP; P02701; 1RAV.
DR   InterPro; IPR005468; Avidin/str.
DR   Pfam; PF01382; Avidin; 1.
DR   PROSITE; PS00577; AVIDIN; 1.
KW   Biotin; Signal; Multigene family.
FT   SIGNAL        1      24       POTENTIAL.
FT   CHAIN        25     150       AVIDIN-RELATED PROTEIN 6.
FT   BINDING      57      57       BIOTIN (NON-COVALENT) (BY SIMILARITY).
FT   CARBOHYD     54      54       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     93      93       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    141     141       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   150 AA;  16528 MW;  C0F3397DF291B150 CRC64;

  Query Match             0.7%;  Score 7;  DB 1;  Length 150;
  Best Local Similarity   100.0%;  Pred. No. 25;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        744 KWLQRLA 750
              |||||||
Db        118 KWLQRLA 124


RESULT 14
AVR7_CHICK
ID   AVR7_CHICK     STANDARD;      PRT;   150 AA.
AC   P56736; Q9W6V4;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Avidin-related protein 7 precursor.
GN   AVR7.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Rhode Island;
RX   MEDLINE=21070478; PubMed=11167523;
RA   Ahlroth M.K., Kola E.H., Ewald D., Masabanda J., Sazanov A., Fries R.,
RA   Kulomaa M.S.;
RT   "Characterization and chromosomal localization of the chicken avidin
RT   gene family.";
RL   Anim. Genet. 31:367-375(2000).
CC   -!- SIMILARITY: BELONGS TO THE AVIDIN/STREPTAVIDIN FAMILY.
CC   ------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to licensewisb-sib.ch).
CC   ------------------------------------------------------------------------
DR   EMBL; AJ237659; CAB39894.1; -.
DR   HSSP; P02701; 1RAV.
DR   InterPro; IPR005468; Avidin/str.
DR   Pfam; PF01382; Avidin; 1.
DR   PROSITE; PS00577; AVIDIN; 1.
KW   Biotin; Signal; Multigene family.
FT   SIGNAL        1      24       POTENTIAL.
FT   CHAIN        25     150       AVIDIN-RELATED PROTEIN 7.
FT   BINDING      57      57       BIOTIN (NON-COVALENT) (BY SIMILARITY).
FT   CARBOHYD     41      41       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     54      54       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     93      93       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    141     141       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   150 AA;  16543 MW;  30F335611F65923D CRC64;

  Query Match             0.7%;  Score 7;  DB 1;  Length 150;
  Best Local Similarity   100.0%;  Pred. No. 25;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        744 KWLQRLA 750
              |||| ||
Db        118 KWLQRLA 124


RESULT 15
DHB_CRAPL
ID   DHB_CRAPL      STANDARD;      PRT;   155 AA.
AC   P22239;
DT   01-AUG-1991 (Rel. 19, Created)
DT   01-AUG-1991 (Rel. 19, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Desiccation-related protein clone PCC6-19 (CDET6-19).
OS   Craterostigma plantagineum
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC   Asteridae; lamiids; Lamiales; Lamiales incertae sedis; Torenieae;
OC   Craterostigma.
OX   NCBI_TaxID=4153;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Leaf;
RA   Piatkowski D., Schneider K., Salamini F., Bartels D.;
RT   "Characterization of five abscisic acid-responsive cDNA clones
RT   isolated from the desiccation-tolerant plant Craterostigma
```

plantagineum and their relationship to other water-stress genes.";
Plant Physiol. 94:1682-1688(1990).
[2]
SEQUENCE FROM N.A.
MEDLINE=94207182; PubMed=8155876;
Michel D., Furini A., Salamini F., Bartels D.;
"Structure and regulation of an ABA- and desiccation-responsive gene
from the resurrection plant Craterostigma plantagineum.";
Plant Mol. Biol. 24:549-560(1994).
-!- INDUCTION: BY DESICCATION (LEAVES) AND BY ABSCISIC ACID (LEAVES
AND CALLUS).
-!- SIMILARITY: BELONGS TO THE PLANT DEHYDRIN FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).

EMBL; M62988; AAA63613.1; -.
EMBL; X74067; CAB59720.1; -.
PIR; S43775; S43775.
InterPro; IPR000167; Dehydrin.
Pfam; PF00257; dehydrin; 1.
PROSITE; PS00315; DEHYDRIN_1; 1.
PROSITE; PS00823; DEHYDRIN_2; 2.
Dehydrin.
DOMAIN 82 93 SER-RICH.
DOMAIN 101 111 LYS-RICH (BASIC).
DOMAIN 139 149 LYS-RICH (BASIC).
SEQUENCE 155 AA; 15568 MW; D3A0813A6141E6EE CRC64;

Query Match 0.7%; Score 7; DB 1; Length 155;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

106 DEYGNPI 112
|||||||
15 DEYGNPI 21

earch completed: October 29, 2003, 18:36:16
ob time : 23.8085 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein protein search, using sw model

Run on: October 29, 2003, 18:42:09 ; Search time 30.2669 Seconds
(without alignments)
3107.451 Million cell updates/sec

Title: US 09-980-043 2
Perfect score: 978
Sequence: 1 MMESSGAPICHTCGEQVGHDA..... .PFVNKV.SILVARTCISIDC 978

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283308 seqs, 96168682 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR 76:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	61	6.2	974	2	T10797	cellulose synthase
2	47	4.8	958	2	T04870	cellulose synthase
3	43	4.4	583	2	T02209	cellulose synthase
4	39	4.0	685	2	T17870	cellulose synthase
5	32	3.3	1081	2	I51318	cellulose synthase
6	32	3.3	1084	2	T08563	cellulose synthase
7	29	3.0	1088	2	B84624	probable cellulose
8	24	2.5	1026	2	T51579	cellulose synthase
9	23	2.4	1081	2	T05361	cellulose synthase
10	22	2.2	1065	2	T52054	cellulose synthase
11	20	2.0	1065	2	B84649	probable cellulose
12	12	1.2	748	2	T02560	cellulose synthase
13	12	1.2	979	2	C86446	probable cellulose
14	12	1.2	1145	2	T51546	cellulose synthase
15	12	1.2	1181	2	D86167	hypothetical prote
16	11	1.1	1036	2	D84741	probable cellulose
17	11	1.1	1111	2	T06646	hypothetical prote
18	9	0.9	285	2	B95943	probable sugar upt
19	9	0.8	192	2	T02854	hypothetical prote
20	8	0.8	326	2	T08591	TGACG-motif bindin
21	8	0.8	344	2	E84377	protein export (im
22	8	0.8	532	2	C96514	hypothetical prote
23	8	0.8	674	2	T52347	hypothetical prote
24	8	0.8	1171	2	E97295	transcription -rep
25	8	0.8	1694	2	S50065	staloadhesin - mou
26	7	0.7	44	2	G64536	hypothetical prote
27	7	0.7	65	2	T15856	hypothetical prote
28	7	0.7	65	4	IMBP14	IMBP14 immun
29	7	0.7	66	2	E70085	hypothetical prote
30	7	0.7	69	1	QQVZ7	hypothetical prote
31	7	0.7	69	2	F42516	D-ORF-C protein -
32	7	0.7	86	2	AD3376	hypothetical prote
33	7	0.7	86	2	AC2117	hypothetical prote
34	7	0.7	92	2	AI3345	hypothetical prote
35	7	0.7	99	2	H70868	hypothetical prote
36	7	0.7	112	2	T25554	hypothetical prote
37	7	0.7	113	1	IMECE1	colicin E1 immunit
38	7	0.7	113	2	I64785	imm protein - Esch
39	7	0.7	116	2	S14211	NADH2 dehydrogenas
40	7	0.7	117	2	S07933	hypothetical prote
41	7	0.7	117	2	G75135	hypothetical prote
42	7	0.7	120	2	S65067	dehydrin DHN1 po
43	7	0.7	125	2	G71000	hypothetical prote
44	7	0.7	132	2	T26625	hypothetical prote
45	7	0.7	132	2	T30595	sugar transport ho

ALIGNMENTS

RESULT 1
T10797
cellulose synthase (EC 2.4.1.-) catalytic chain celA1 - upland cotton
C;Species: Gossypium hirsutum (upland cotton)
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
C;Accession: T10797
R;Pear, J.R.; Kawagoe, Y.; Schreckengost, W.E.; Delmer, D.P.; Stalker, D.M.
Proc. Natl. Acad. Sci. U.S.A. 93, 12637-12642, 1996
A;Title: Higher plants contain homologs of the bacterial celA genes encoding the cata
A;Reference number: Z17152; MUID:97057296; PMID:8901635
A;Accession: T10797
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-974 <PEA>
A;Cross-references: EMBL:U58283; NID:g1706955; PIDN:AAB37766.1; PID:g1706956
A;Experimental source: strain Acala SJ-2; fiber
C;Genetics:
A;Gene: celA1
C;Function:
A;Description: involved in the synthesis of cellulose
C;Keywords: glycosyltransferase; hexosyltransferase

Query Match 6.2%; Score 61; DB 2; Length 974;
Best Local Similarity 100.0%; Pred. No. 3.1e-55;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy        798 SIIVTAVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQGFLKMLAGIDTNFTVTAKA 858
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        794 SIIVTAVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQGFLKMLAGIDTNFTVTAKA 853

Qy        859 A 859
              |
Db        854 A 854
```

RESULT 2
T04870
cellulose synthase (EC 2.4.1.-) catalytic chain F28A21.190 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-Jun-1999
C;Accession: T04870
R;Bevan, M.; Mueller, M.W.; Muendlein, A.; Felber, R.; Bancroft, I.; Mewes, H.W.; May
submitted to the Protein Sequence Database, February 1999
A;Reference number: Z15387
A;Accession: T04870
A;Molecule type: DNA
A;Residues: 1-958 <BEV>
A;Cross-references: EMBL:AL035526
A;Experimental source: cultivar Columbia; BAC clone F28A21
C;Genetics:
A;Map position: 4
A;Introns: 60/1; 76/3; 94/1; 149/2; 240/2; 355/3; 401/3; 443/3; 514/3; 662/3; 780/3;

Note: F28A21.190
Keywords: glycosyltransferase; hexosyltransferase

Query Match 4.8%; Score 47; DB 2; Length 958;
Best Local Similarity 100.0%; Pred. No. 1.9e-40;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

409 GARDIEGNELPRLVYVSREKRPGYQHHKKAGAENALVRVSAVLTNAP 455
408 GARDIEGNELPRLVYVSREKRPGYQHHKKAGAENALVRVSAVLTNAP 454

SULT 3
2209
llulose synthase (EC 2.4.1.-) catalytic chain - rice (fragment)
Species: Oryza sativa (rice)
Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 11-Jun-1999
Accession: T02209
Arioli, T.; Peng, L.; Betzner, A.S.; Burn, J.; Wittke, W.; Herth, W.; Camilleri, C.; H
ience 279, 717-720, 1998
Title: Molecular analysis of cellulose biosynthesis in Arabidopsis.
Reference number: Z13745; MUID:98111412; PMID:9445479
Accession: T02209
Status: translated from GB/EMBL/DDBJ
Molecule type: mRNA
Residues: 1-583 <ARI>
Cross-references: EMBL:AF030052; NID:g2781432; PID:g2781433
Experimental source: subsp. Japonica, strain Nipponbare
Keywords: glycosyltransferase; hexosyltransferase

Query Match 4.1%; Score 40; DB 2; Length 583;
Best Local Similarity 100.0%; Pred. No. 3.1e-33;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

256 FVSTVDPLKEPPLITANTVLSILAVDYPVDKVSCYVSDDG 295
354 FVSTVDPLKEPPLITANTVLSILAVDYPVDKVSCYVSDDG 393

CSULT 4
0800
ellulose synthase (EC 2.4.1.-) catalytic chain celA2 - upland cotton (fragment)
Species: Gossypium hirsutum (upland cotton)
,Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
,Accession: T10800
,Pear, J.R.; Kawagoe, Y.; Schreckengost, W.E.; Delmer, D.P.; Stalker, D.M.
roc. Natl. Acad. Sci. U.S.A. 93, 12637-12642, 1996
,Title: Higher plants contain homologs of the bacterial celA genes encoding the catalyt
,Reference number: Z17152; MUID:97057296; PMID:8901635
,Accession: T10800
,Status: preliminary; translated from GB/EMBL/DDBJ
;Molecule type: mRNA
;Residues: 1-685 <PEA>
;Cross-references: EMBL:U58284; NID:g1706957; PIDN:AAB37767.1; PID:g1706958
;Experimental source: strain Acala SJ-2; fiber
;Genetics:
;Gene: celA2
;Function:
;Description: involved in synthesis of cellulose
;Keywords: glycosyltransferase; hexosyltransferase

Query Match 4.0%; Score 39; DB 2; Length 685;
Best Local Similarity 100.0%; Pred. No. 4e-32;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

y 417 ELPRLVYVSREKRPGYQHHKKAGAENALVRVSAVLTNAP 455
b 105 ELPRLVYVSREKRPGYQHHKKAGAENALVRVSAVLTNAP 143

RESULT 5
T52028
cellulose synthase [imported] - Arabidopsis thaliana (fragment)
C;Species: Arabidopsis thaliana (mouse ear cress)
C;Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 20-Oct-2000
C;Accession: T52028
R;Joshi, C.
submitted to the EMBL Data Library, May 1998
A;Reference number: Z25890
A;Accession: T52028
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1081 <JOS>
A;Cross-references: EMBL:AF[illegible]; PIDN:AAC39367.1

Query Match [illegible]; Score 32; DB 2; Length 1081;
Best Local Similarity [illegible]; Pred. No. 1.6e-24;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 256 FVSTVDPLKEPPLI[illegible] 287
Db 355 FVSTVD[illegible] 386

RESULT 6
T08583
cellulose synthase (EC 2.4.1.-) catalytic chain - Arabidopsis thaliana
N;Alternate names: protein T22F8.250
C;Species: Arabidopsis thaliana (mouse ear cress)
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 22-Oct-1999
C;Accession: T08583; T09014
R;Bevan, M.; Zimmermann, W.; Bruneisen, A.; Wambutt, R.; Bancroft, I.; Mewes, H.W.; May
submitted to the Protein Sequence Database, May 1999
A;Reference number: Z16442
A;Accession: T08583
A;Molecule type: DNA
A;Residues: 1-1084 <BEV>
A;Cross-references: EMBL:AL[illegible]; GSPDB:GN00[illegible]; ATSP:T22F8.250
A;Experimental source: cultivar Columbia; BAC clone T22F8
R;Arioli, T.; Peng, L.; Betzner, A.S.; Burn, J.; Wittke, W.; Herth, W.; Camilleri, C.; H
Science 279, 717-720, 1998
A;Title: Molecular analysis of cellulose biosynthesis in Arabidopsis
A;Reference number: Z13745; MUID:98111412; PMID:9445479
A;Accession: T09014
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1084 <ARI>
A;Cross-references: EMBL:AF027173; NID:g2827140; PIDN:AAC39335.1; PID:g2827141
A;Experimental source: cultivar Columbia
C;Genetics:
A;Gene: ATSP:T22F8.250; Ath A
A;Map position: 4
A;Introns: 27/3; 156/3; 191/1; 254/2; 343/2; 456/3; 504/3; 546/3; 617/3; 708/1; 773/3; #
C;Keywords: cell wall synthesis; glycosyltransferase; hexosyltransferase

Query Match [illegible]; Score 32; DB 2; Length 1084;
Best Local Similarity [illegible]; Pred. No. 1.6e-24;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 256 FVSTVDPLKEPPLI[illegible] 287
Db 356 FVSTVDPLKEPPLI[illegible] 387

RESULT 7
H84604
probable cellulose synthase catalytic subunit [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: H84604
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197

A;Accession: H84604
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1 1098 <STO>
A;Cross-references: GB:AE002093; NID:g4417271; PIDN:AAD20396.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g21770
A;Map position: 2

Query Match 3.0%; Score 24; DB 2; Length 1098;
Best Local Similarity 100.0%; Pred. No. 2.2e-21;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 256 FVSTVDPLKEPPLITANTVLSILAVDYPV 284
||||||||||||||||||||||||||||
Db 369 FVSTVDPLKEPPLITANTVLSILAVDYPV 392

RESULT 8
T51579
cellulose synthase catalytic subunit (IRX9) - Arabidopsis thaliana
N;Alternate names: protein T10B6_80
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 18-Aug-2000
C;Accession: T51579
R;Sato, S.; Nakamura, Y.; Kaneko, T.; Kato, T.; Asamizu, E.; Kotani, H.; Tabata, S.; Mew
submitted to the Protein Sequence Database, August 2000
A;Reference number: Z25394
A;Accession: T51579
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1026 <SAT>
A;Cross-references: EMBL:AL391142
A;Experimental source: cultivar Columbia; BAC clone T10B6
C;Genetics:
A;Map position: 5
A;Introns: 25/3; 91/1; 184/1; 215/2; 304/2; 419/3; 507/3; 578/3; 650/1; 715/3; 833/3
A;Note: T10B6_80

Query Match 2.5%; Score 24; DB 2; Length 1026;
Best Local Similarity 100.0%; Pred. No. 4e-16;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 306 ETARRARKWVPFCKKFSIEPRAFF 329
||||||||||||||||||||||||
Db 870 ETARRARKWVPFCKKFSIEPRAFF 893

RESULT 9
T05351
cellulose synthase (EC 2.4.1.-) catalytic chain RSW1 - Arabidopsis thaliana
N;Alternate names: protein F8B4.110
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-Jun-1999
C;Accession: T05351
R;Bevan, M.; Terryn, N.; Ardiles, W.; Buysshaert, C.; Rousseville, R.; De Clerck, R.; De
ewes, H.W.; Mayer, K.F.X.; Schueller, C.
submitted to the Protein Sequence Database, February 1999
A;Reference number: Z15409
A;Accession: T05351
A;Molecule type: DNA
A;Residues: 1-1081 <BEV>
A;Cross-references: EMBL:AL034567
A;Experimental source: cultivar Columbia; BAC clone F8B4
C;Genetics:
A;Gene: RSW1
A;Map position: 4
A;Introns: 27/3; 93/1; 150/3; 189/1; 262/2; 441/2; 456/1; 502/3; 544/3; 615/3; 704/1; 76
A;Note: F8B4.110
C;Keywords: glycosyltransferase; hexosyltransferase; P-loop

Query Match 2.4%; Score 23; DB 2; Length 1081;
Best Local Similarity 100.0%; Pred. No. 4.7e-19;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 376 PEEGWTMQDGTPWPGNNTRDHPG 398
|||||||||||||||||||||||
Db 477 PEEGWTMQDGTPWPGNNTRDHPG 499

RESULT 10
T52054
cellulose synthase (EC 2.4.1.-) catalytic subunit (validated) - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 31-Dec-2000
C;Accession: T52054
R;Arioli, T.; Peng, L.; Betzner, A.S.; Burn, J.; Wittke, W.; Herth, W.; Camilleri, C.
Science 279, 717-720, 1998
A;Title: Molecular analysis of cellulose biosynthesis in Arabidopsis.
A;Reference number: Z13745; MUID:98111412; PMID:9445479
A;Accession: T52054
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1065 <ARI>
A;Cross-references: EMBL:AF027174; PIDN:AAC39336.1
C;Genetics:
A;Note: Ath-B
C;Function:
A;Description: EC 2.4.1.-; cellulose synthase [validated, MUID:98111412]; involved in
C;Keywords: glycosyltransferase; hexosyltransferase

Query Match 2.2%; Score 22; DB 2; Length 1065;
Best Local Similarity 100.0%; Pred. No. 5.3e-14;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 270 TANTVLSILAVDYPVDKVSCYV 291
||||||||||||||||||||||
Db 355 TANTVLSILAVDYPVDKVSCYV 376

RESULT 11
F84649
probable cellulose synthase catalytic subunit [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: F84649
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: F84649
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1065 <STO>
A;Cross-references: GB:AE002093; NID:g4432865; PIDN:AAD20713.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g25540
A;Map position: 2

Query Match 2.0%; Score 20; DB 2; Length 1065;
Best Local Similarity 100.0%; Pred. No. 6.9e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 323 IEPRAPEFYFSQKIDYLKDK 342
||||||||||||||||||||
Db 411 IEPRAPEFYFSQKIDYLKDK 430

RESULT 12
T02560
cellulose synthase homolog T26B15.17 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 16-Feb-2001
C;Accession: T02560; C84735

Rounsley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes,
bmitted to the EMBL Data Library, July 1998
Description: Arabidopsis thaliana chromosome II BAC T26B15 genomic sequence.
Reference number: Z14678
Accession: T02560
Status: translated from GB/EMBL/DDBJ
Molecule type: DNA
Residues: 1-748 <ROU>
Cross-references: EMBL:AC004681; NID:g3298532; PID:g3298549
Experimental source: cultivar Columbia
Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
ss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
ture 402, 761-768, 1999
Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
Reference number: A84420; MUID:20083487; PMID:10617197
Accession: C84735
Status: preliminary
Molecule type: DNA
Residues: 1-748 <STO>
Cross-references: GB:AE002093; NID:g3298549; PIDN:AAC25943.1; GSPDB:GN00139
Genetics:
Gene: At2g32610; T26B15.17
Map position: 2
Introns: 88/2; 195/3; 237/3; 278/3; 345/3; 391/1; 440/1; 558/3

Query Match 1.2%; Score 12; DB 2; Length 748;
Best Local Similarity 100.0%; Pred. No. 0.0014;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

y 452 TNAPYILNVDCD 463
||||||||||||
b 286 TNAPYILNVDCD 297

ESULT 13
86446
robable cellulose synthase catalytic subunit [imported] - Arabidopsis thaliana
;Species: Arabidopsis thaliana (mouse-ear cress)
;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
;Accession: C86446
;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
nsen, N.F.; Hughes, B.; Huizar, L.
ature 408, 816-820, 2000
;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
izzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
er, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: C86446
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-979 <STO>
A;Cross-references: GB:AE005172; NID:g10801364; PIDN:AAG23436.1; GSPDB:GN00141
C;Genetics:
A;Map position: 1

Query Match 1.2%; Score 12; DB 2; Length 979;
Best Local Similarity 100.0%; Pred. No. 0.0018;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 421 LVYVSREKRPGY 432
||||||||||||
Db 431 LVYVSREKRPGY 442

RESULT 14
T51546
cellulose synthase catalytic subunit-like protein - Arabidopsis thaliana
N;Alternate names: protein F2K13_60
C;Species: Arabidopsis tha [illegible]
C;Date: 18-Aug-2000 #se[illegible] revision 18-Aug-2000 #text_change 18-Aug-2000
C;Accession: T51546
R;Sato, S.; Nakamura, Y.; Ka[illegible]; Sato, T.; Asamizu, E.; Kotani, H.; Tabata, S.; You
submitted to the Protein Sequence Database, August 2000
A;Reference number: Z25434
A;Accession: T51546
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1148 <SAT>
A;Cross-references: EMBL:A[illegible]
A;Experimental source: mul[illegible]; BAC clone F2K13
C;Genetics:
A;Map position: 5
A;Introns: 397/2; 566/3
A;Note: F2K13_60

Query Match [illegible] Length 1148;
Best Local Similarity [illegible]
Matches 12; Conservative [illegible] Mismatches [illegible] Indels 0; Gaps 0;

Qy 421 LVYVSREKRPGY [illegible]
||.|||.:||
Db 567 LVYVSREKRPGY [illegible]

RESULT 15
D86157
hypothetical protein F22D16.6 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Dec-2001
C;Accession: D86157
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: D86157
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1181 <STO>
A;Cross-references: GB:AE005172; NID:g6056428; PIDN:AAF02892.1; GSPDB:GN00141
C;Genetics:
A;Map position: 1

Query Match 1.2%; Score 12; DB 2; Length 1181;
Best Local Similarity [illegible]; Pred. No. [illegible];
Matches 12; Conservative [illegible] Mismatches 0; Indels 0; Gaps 0;

Qy 421 LVYVSREKRPGY 432
|||.|||.:||
Db 622 LVYVSREKRPGY 633

Search completed: October [illegible], 16:36:56
Job time : 33.2669 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame plus_p2n model

Run on: October 31, 2003, 23:56:54 ; Search time 747.215 Seconds
(without alignments)
2559.830 Million cell updates/sec

Title: US-09-980-043-2
Perfect score: 978
Sequence: 1 MMESGAPICHTCGEQVTDA...PEVMEVTILVASCISICC 978

Scoring table: OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1811591 seqs, 1355896290 residues

Word size: 3

Total number of hits satisfying chosen parameters: 3617486

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post processing: Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09980043/runat_29102003_163934_22798/app_query.fasta_1.1358
-DB=Published_Applications_NA -QFMT=fastap -SUFFIX=oligo -rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=oligo
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1
-ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09980043_gCGN_1_1_350_runat_29102003_163934_22798
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSTBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US08_PUB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US09A_PUB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US09A_PUB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	61	6.2	3328	10	US-09-838-539-1	Sequence 1, Appli
2	41	4.2	3231	12	US-10-209-059-27	Sequence 27, Appl
3	40	4.1	1741	14	US-10-229-193-13	Sequence 13, Appl
4	40	4.1	3725	12	US-10-209-059-45	Sequence 45, Appl
5	40	4.1	3725	12	US-10-160-719-13	Sequence 13, Appl
6	40	4.1	3725	12	US-10-160-719-21	Sequence 21, Appl
7	40	4.1	3725	12	US-10-160-719-41	Sequence 41, Appl
8	36	3.7	2890	10	US-09-900-237-17	Sequence 17, Appl
9	35	3.6	3614	14	US-10-229-193-9	Sequence 9, Appli
10	32	3.3	3255	10	US-09-938-842A-559	Sequence 559, App
11	32	3.3	3255	10	US-09-938-842A-2530	Sequence 2530, Ap
12	32	3.3	3828	14	US-10-229-193-7	Sequence 7, Appli
13	32	3.3	5009	14	US-10-229-193-4	Sequence 4, Appli
14	30	3.1	1189	10	US-09-900-237-5	Sequence 5, Appli
15	30	3.1	2663	10	US-09-900-237-23	Sequence 23, Appl
16	30	3.1	3470	12	US-10-209-059-25	Sequence 25, Appl
17	30	3.1	3704	12	US-10-209-059-13	Sequence 13, Appl
18	30	3.1	3704	12	US-10-160-719-57	Sequence 57, Appl
19	30	3.1	3746	12	US-10-160-719-29	Sequence 29, Appl
20	30	3.1	3746	12	US-10-160-719-49	Sequence 49, Appl
21	30	3.1	3773	12	US-10-209-059-9	Sequence 9, Appli
22	30	3.1	3773	12	US-10-160-719-5	Sequence 5, Appli
23	30	3.1	3799	12	US-10-267-459-5	Sequence 5, Appli
24	30	3.1	3799	12	US-10-209-059-21	Sequence 21, Appl
25	27	2.8	3028	12	US-10-209-059-29	Sequence 29, Appl
26	27	2.8	3626	10	US-09-900-237-29	Sequence 29, Appl
27	26	2.7	3563	10	US-09-900-237-25	Sequence 25, Appl
28	24	2.5	2830	12	US-10-267-459-1	Sequence 1, Appli
29	24	2.5	2830	12	US-10-209-059-5	Sequence 5, Appli
30	24	2.5	3517	10	US-09-900-237-13	Sequence 13, Appl
31	23	2.4	3603	14	US-10-229-193-5	Sequence 5, Appli
32	23	2.4	3673	14	US-10-229-193-11	Sequence 11, Appl
33	23	2.4	3753	12	US-10-160-719-33	Sequence 33, Appl
34	23	2.4	3753	12	US-10-160-719-53	Sequence 53, Appl
35	23	2.4	3780	12	US-10-209-059-1	Sequence 1, Appli
36	23	2.4	3780	12	US-10-160-719-9	Sequence 9, Appli
37	23	2.4	4612	10	US-09-838-539-2	Sequence 2, Appli
38	23	2.4	8411	14	US-10-229-193-3	Sequence 3, Appli
39	22	2.2	258	9	US-09-923-876-2317	Sequence 2317, Ap
40	22	2.2	3786	10	US-09-900-237-7	Sequence 7, Appli
41	22	2.2	3813	12	US-10-209-059-17	Sequence 17, Appl
42	22	2.2	3813	12	US-10-160-719-25	Sequence 25, Appl
43	22	2.2	3813	12	US-10-160-719-45	Sequence 45, Appl
44	21	2.1	2248	14	US-10-229-193-1	Sequence 1, Appli
45	20	2.0	1560	10	US-09-900-237-27	Sequence 27, Appl

ALIGNMENTS

RESULT 1
US-09-838-539-1
; Sequence 1, Application US/09838539
; Patent No. US20020129401A1
; GENERAL INFORMATION:
; APPLICANT: Stalker, D. et al.
; TITLE OF INVENTION: Plant Cellulose Synthase and Promoter
; TITLE OF INVENTION: Sequences
; FILE REFERENCE: 15621/03/US
; CURRENT APPLICATION NUMBER: US/09/838,539
; CURRENT FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/029,987
; PRIOR FILING DATE: 1996-10-29
; PRIOR APPLICATION NUMBER: 08/960,048
; PRIOR FILING DATE: 1997-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 3328
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Oligonucleotide
US-09-838-539-1

gnment Scores:
ed. No.: 5.82e-54 Length: 3328
ore: 61.00 Matches: 61
rcent Similarity: 100.00% Conservative: 0
st Local Similarity: 100.00% Mismatches: 0
ery Match: 6.24% Indels: 0
10 Gaps: 0

-09-980-043-2 (1-978) x US-09-838-539-1 (1-3328)

799 SerIleIleValThrAlaValLeuGluLeuArgTrpSerGlyValSerIleGluAspLeu 818
2558 TCCATTATCGTGACTGCTGTTCTCGAGCTCCGATGGAGTGGTGTCAGCATTGAGGACTTA 2617

819 TrpArgAsnGluGlnPheTrpValIleGlyGlyValSerAlaHisLeuPheAlaValPhe 838
2618 TGGCGTAACGAGCAGTTTTGGGTCATCGGTGGCGTTTCAGCCCATCTCTTTGCCGTCTTC 2677

839 GlnGlyPheLeuLysMetLeuAlaGlyIleAspThrAsnPheThrValThrAlaLysAla 858
2678 CAAGGTTTCCTTAAGATGCTTGCGGGCATTGACACCAACTTTACTGTCACTGCCAAAGCA 2737

859 Ala 859
2738 GCT 2740

SULT 2
-10-209-059-27
Sequence 27, Application US/10209059
Publication No. US20030163838A1
GENERAL INFORMATION:
APPLICANT: Dhugga, Kanwarpal S.
APPLICANT: Wang, Haiyin
TITLE OF INVENTION: Maize Cellulose Synthases and Uses
TITLE OF INVENTION: Thereof
FILE REFERENCE: 0864R2
CURRENT APPLICATION NUMBER: US/10/209,059
CURRENT FILING DATE: 2002-07-31
PRIOR APPLICATION NUMBER: 60/096,822
PRIOR FILING DATE: 1998-08-17
PRIOR APPLICATION NUMBER: 09/371,383
PRIOR FILING DATE: 1999-08-06
PRIOR APPLICATION NUMBER: 09/550,483
PRIOR FILING DATE: 2000-04-14
NUMBER OF SEQ ID NOS: 52
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 27
LENGTH: 3231
TYPE: DNA
ORGANISM: Zea mays
S-10-209-059-27

lignment Scores:
red. No.: 9e-33 Length: 3231
core: 41.00 Matches: 41
ercent Similarity: 100.00% Conservative: 0
est Local Similarity: 100.00% Mismatches: 0
uery Match: 4.19% Indels: 0
B: 12 Gaps: 0

S-09-980-043-2 (1-978) x US-10-209-059-27 (1-3231)

y 681 GlyPheLysMetHisCysArgGlyTrpArgSerIleTyrCysMetProValArgProAla 700
b 2076 GGGTTTAAGATGCACTGCCGTGGCTGGAGATCCATCTACTGCATGCCGGTGAGACCTGCA 2135

y 701 PheLysGlySerAlaProIleAsnLeuSerAspArgLeuHisGlnValLeuArgTrpAla 720
b 2136 TTCAAGGGATCAGCCCCAATCAATCTTTCCGATCGTCTTCACCAAGTTCTCCGGTGGGCT 2195

)y 721 Leu 721
)b 2196 CTT 2198

RESULT 3
US-10-229-193-13
; Sequence 13, Application [illegible]
; Publication No. US20030126643A1
; GENERAL INFORMATION:
; APPLICANT: Arioli, Antoni Williamson, Richard E.
; APPLICANT: Betzner, Andreas S.
; APPLICANT: Peng, Liangcai
; TITLE OF INVENTION: Manipulation of Cellulose and/or Beta-1,4-Glucan
; FILE REFERENCE: 96-98A
; CURRENT APPLICATION NUMBER: US/10/229,193
; CURRENT FILING DATE: 2002-08-26
; PRIOR APPLICATION NUMBER: 09/271,314
; PRIOR FILING DATE: 1998-[illegible]
; PRIOR APPLICATION NUMBER: [illegible]
; PRIOR FILING DATE: 1999-[illegible]
; PRIOR APPLICATION NUMBER: [illegible]
; PRIOR FILING DATE: 1998-[illegible]
; NUMBER OF SEQ ID NOS: [illegible]
; SOFTWARE: PatentIn [illegible]
; SEQ ID NO 13
; LENGTH: 1741
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (101)..(1741)
US-10-229-193-13

Alignment Scores:
Pred. No.: 5.74e-[illegible] Length: 1741
Score: 40.2 Matches: 40
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 4.00% Indels: 0
DB: 14 Gaps: 0

US-09-980-043-2 (1-978) x US-10-229-193-13 (1-1741)

Qy 256 PheValSerThrValAspProLeuLysValProProLeuIleThrAlaAsnThrValLeu 275
Db 1160 TTTGTCAGTACGGTAGATCCACTAAAAGTACCTCCTCTGATCACAGCAAACACTGTTTTG 1219

Qy 276 SerIleLeuAlaValAspTyrProValAspLysValSerCysTyrValSerAspAspGly 295
Db 1220 TCCATTCTGGCTGTTATTACTCTGTTGACAAAGTGTCATGCTATGTTTCTGACGATGGT 1279

RESULT 4
US-10-209-059-45
; Sequence 45, Application US/10209059
; Publication No. US20030163838A1
; GENERAL INFORMATION:
; APPLICANT: Dhugga, Kanwarpal S.
; APPLICANT: Wang, Haiyin
; TITLE OF INVENTION: Maize Cellulose Synthases and Uses
; TITLE OF INVENTION: Thereof
; FILE REFERENCE: 0864R2
; CURRENT APPLICATION NUMBER: US/10/209,059
; CURRENT FILING DATE: 2002-07-31
; PRIOR APPLICATION NUMBER: 60/096,822
; PRIOR FILING DATE: 1998-08-17
; PRIOR APPLICATION NUMBER: 09/371,383
; PRIOR FILING DATE: 1999-08-06
; PRIOR APPLICATION NUMBER: 09/550,483
; PRIOR FILING DATE: 2000-04-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 45
; LENGTH: 3725
; TYPE: DNA
; ORGANISM: Zea mays
US-10-209-059-45

Alignment Scores:
Pred. No.: 1.18e-31 Length: 3725
Score: 40.00 Matches: 40
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.09% Indels: 0
DB: 12 Gaps: 0

US-09-980-043-2 (1-978) x US-10-209-719-46 (1-3725)

Qy 256 PheValSerThrValAspProLeuLysGluProProLeuIleThrAlaAsnThrValLeu 275
Db 1232 TTTGTCAGTACAGTGGATCCATTGAAGGAACCTCCACTGATCACAGCCAACACTGTTTTG 1291

Qy 276 SerIleLeuAlaValAspTyrProValAspLysValSerCysTyrValSerAspAspGly 295
Db 1292 TCCATTCTTGCTGTGGATTACCCTGTTGACAAAGTGTCATGCTATGTTTCTGATGATGGC 1351

RESULT 5
US-10-160-719-13
; Sequence 13, Application US/10160719
; Publication No. US20030167528A1
; GENERAL INFORMATION:
; APPLICANT: Kanwarpal S. Dhugga
; APPLICANT: Timothy G. Helentjaris
; APPLICANT: Benjamin A. Bowen
; APPLICANT: Xun Wang
; TITLE OF INVENTION: Maize Cellulose Synthases and Uses
; TITLE OF INVENTION: Thereof
; FILE REFERENCE: 0864
; CURRENT APPLICATION NUMBER: US/10/160,719
; CURRENT FILING DATE: 2002-06-03
; PRIOR APPLICATION NUMBER: US/09/371,383A
; PRIOR FILING DATE: 1999-08-06
; PRIOR APPLICATION NUMBER: US 60/096,822
; PRIOR FILING DATE: 1998-08-17
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 13
; LENGTH: 3725
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (179)...(3400)
US-10-160-719-13

Alignment Scores:
Pred. No.: 1.18e-31 Length: 3725
Score: 40.00 Matches: 40
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.09% Indels: 0
DB: 12 Gaps: 0

US-09-980-043-2 (1-978) x US-10-160-719-13 (1-3725)

Qy 256 PheValSerThrValAspProLeuLysGluProProLeuIleThrAlaAsnThrValLeu 275
Db 1232 TTTGTCAGTACAGTGGATCCATTGAAGGAACCTCCACTGATCACAGCCAACACTGTTTTG 1291

Qy 276 SerIleLeuAlaValAspTyrProValAspLysValSerCysTyrValSerAspAspGly 295
Db 1292 TCCATTCTTGCTGTGGATTACCCTGTTGACAAAGTGTCATGCTATGTTTCTGATGATGGC 1351

RESULT 6
US-10-160-719-21
; Sequence 21, Application US/10160719
; Publication No. US20030167528A1
; GENERAL INFORMATION:
; APPLICANT: Kanwarpal S. Dhugga
; APPLICANT: Timothy G. Helentjaris
; APPLICANT: Benjamin A. Bowen
; APPLICANT: Xun Wang
; TITLE OF INVENTION: Maize Cellulose Synthases and Uses
; TITLE OF INVENTION: Thereof
; FILE REFERENCE: 0864
; CURRENT APPLICATION NUMBER: US/10/160,719
; CURRENT FILING DATE: 2002-06-03
; PRIOR APPLICATION NUMBER: US/09/371,383A
; PRIOR FILING DATE: 1999-08-06
; PRIOR APPLICATION NUMBER: US 60/096,822
; PRIOR FILING DATE: 1998-08-17
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 21
; LENGTH: 3725
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (179)...(3400)
US-10-160-719-21

Alignment Scores:
Pred. No.: 1.18e-31 Length: 3725
Score: 40.00 Matches: 40
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.09% Indels: 0
DB: 12 Gaps: 0

US-09-980-043-2 (1-978) x US-10-160-719-21 (1-3725)

Qy 256 PheValSerThrValAspProLeuLysGluProProLeuIleThrAlaAsnThrValLeu 275
Db 1232 TTTGTCAGTACAGTGGATCCATTGAAGGAACCTCCACTGATCACAGCCAACACTGTTTTG 1291

Qy 276 SerIleLeuAlaValAspTyrProValAspLysValSerCysTyrValSerAspAspGly 295
Db 1292 TCCATTCTTGCTGTGGATTACCCTGTTGACAAAGTGTCATGCTATGTTTCTGATGATGGC 1351

RESULT 7
US-10-160-719-41
; Sequence 41, Application US/10160719
; Publication No. US20030167528A1
; GENERAL INFORMATION:
; APPLICANT: Kanwarpal S. Dhugga
; APPLICANT: Timothy G. Helentjaris
; APPLICANT: Benjamin A. Bowen
; APPLICANT: Xun Wang
; TITLE OF INVENTION: Maize Cellulose Synthases and Uses
; TITLE OF INVENTION: Thereof
; FILE REFERENCE: 0864
; CURRENT APPLICATION NUMBER: US/10/160,719
; CURRENT FILING DATE: 2002-06-03
; PRIOR APPLICATION NUMBER: US/09/371,383A
; PRIOR FILING DATE: 1999-08-06
; PRIOR APPLICATION NUMBER: US 60/096,822
; PRIOR FILING DATE: 1998-08-17
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 41
; LENGTH: 3725
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (179)...(3400)
US-10-160-719-41

Alignment Scores:
Pred. No.: 1.18e-31 Length: 3725
Score: 40.00 Matches: 40
Percent Similarity: 100.00% Conservative: 0

```
st Local Similarity:  100.00%        Mismatches:    0
ery Match:            4.09%          Indels:        0
                      12             Gaps:          0

-09-980-043-2 (1-978) x US-10-160-719-41 (1-3725)

        256 PheValSerThrValAspProLeuLysGluProProLeuIleThrAlaAsnThrValLeu 275
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
       1232 TTTGTCAGTACAGTGGATCCATTGAAGGAACCTCCACTGATCACAGCCAACACTGTTTTG 1291

        276 SerIleLeuAlaValAspTyrProValAspLysValSerCysTyrValSerAspAspGly 295
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
       1292 TCCATTCTTGCTGTGGATTACCCTGTTGACAAAGTGTCATGCTATGTTTCTGATGATGGC 1351

SULT 8
-09-900-237-17
Sequence 17, Application US/09900237
Patent No. US20020120124A1
GENERAL INFORMATION:
 APPLICANT: Allen, Stephen
 TITLE OF INVENTION: Plant Cellulose Synthases
 FILE REFERENCE: BB1170 US CIP
 CURRENT APPLICATION NUMBER: US/09/900,237
 CURRENT FILING DATE:  2001-07-06
 PRIOR APPLICATION NUMBER: 60/092,844
 PRIOR FILING DATE: 1998-07-14
 PRIOR APPLICATION NUMBER: PCT/US99/15871
 PRIOR FILING DATE: 1999-07-13
 PRIOR APPLICATION NUMBER: 09/720383
 PRIOR FILING DATE: 2000-12-21
 NUMBER OF SEQ ID NOS: 33
 SOFTWARE: Microsoft Office 97
SEQ ID NO 17
  LENGTH: 2890
  TYPE: DNA
  ORGANISM: Glycine max
3-09-900-237-17

lignment Scores:
red. No.:             1.62e-27       Length:        2890
core:                 36.00          Matches:       36
ercent Similarity:    100.00%        Conservative:  0
est Local Similarity: 100.00%        Mismatches:    0
uery Match:           3.68%          Indels:        0
B:                    10             Gaps:          0

S-09-980-043-2 (1-978) x US-09-900-237-17 (1-2890)

y         698 ArgProAlaPheLysGlySerAlaProIleAsnLeuSerAspArgLeuHisGlnValLeu 717
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
b        1538 CGGCCTGCATTTAAGGGGTCTGCGCCTATCAACCTTTCAGATCGTCTGCACCAAGTTCTT 1597

y         718 ArgTrpAlaLeuGlySerValGluIlePhePheSerArgHisCysPro 733
              ||||||||||||||||||||||||||||||||||||||||||||||||||
b        1598 CGGTGGGCTCTTGGGTCTGTTGAGATTTTTTTCAGCAGACATTGTCCA 1645

ESULT 9
S-10-229-193-9
 Sequence 9, Application US/10229193
 Publication No. US20030126643A1
 GENERAL INFORMATION:
  APPLICANT: Arioli, Antonio Williamson, Richard E.
  APPLICANT:  Betzner, Andreas S.
  APPLICANT:  Peng, Liangcai
  TITLE OF INVENTION: Manipulation of Cellulose and/or Beta-1.4-Glucan
  FILE REFERENCE: 96-98A
  CURRENT APPLICATION NUMBER: US/10/229,193
  CURRENT FILING DATE:  2002-08-26
  PRIOR APPLICATION NUMBER: 09/221,013
  PRIOR FILING DATE: 1998-12-23
  PRIOR APPLICATION NUMBER: PCT/AU97/00402
  PRIOR FILING DATE: 1997-06-24
  PRIOR APPLICATION NUMBER: AU PO0699
;  PRIOR FILING DATE: 1996-[illegible]-27
;  NUMBER OF SEQ ID NOS: 36
;  SOFTWARE: PatentIn Ver[illegible]
; SEQ ID NO 9
;   LENGTH: 3614
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (217)..(341)
US-10-229-193-9

Alignment Scores:
Pred. No.:            2.[illegible]26       Length:        3614
Score:                36.[illegible]        Matches:       35
Percent Similarity:   100.00%               Conservative:  0
Best Local Similarity: 100.00%              Mismatches:    0
Query Match:          3.[illegible]         Indels:        1
DB:                   14                    Gaps:          0

US-09-980-043-2 (1-978) x [illegible] (1-3614)

Qy        290 ThrAlaAsnThrVa[illegible]SerIleLeuAlaValAspTyrProValAspLysValSerCys 289
              ||||||||||||||[illegible]||||||||||||||||||||||||||||||||||||||||||||
Db       1279 ACAGCCAACACAGT[illegible]TCTCTATTCTGGTTGTTACTACCCAGTTGACAAGGTGTCCTGT 1338

Qy        290 TyrValSerAspAspGlyAlaAlaMetLeuSerPheGluSerLeu 304
              ||||||||||||||[illegible]||||||||||||||||||||||
Db       1339 TATGTTTCTGATGA[illegible]GGTGCTGCTATGTTATCATTTGAATCACTT 1393

RESULT 10
US-09-938-842A-559
; Sequence 559, Application[illegible]S/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
;  APPLICANT: Harper, Jeff
;  APPLICANT:  Kreps, Joel
;  APPLICANT:  Wang, Xun
;  APPLICANT:  Zhu, Tong
;  TITLE OF INVENTION: STRESS REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
;  TITLE OF INVENTION:  SAME AND METHODS OF USE
;  FILE REFERENCE: SCRIP1300.3
;  CURRENT APPLICATION NUMBER: US/09/938,842A
;  CURRENT FILING DATE:  2001-08-24
;  PRIOR APPLICATION NUMBER: US 60/227,866
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: US 60/264,647
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/[illegible],111
;  PRIOR FILING DATE: 2001-[illegible]-22
;  NUMBER OF SEQ ID NOS: 5[illegible]
; SEQ ID NO 559
;   LENGTH: 3255
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
US-09-938-842A-559

Alignment Scores:
Pred. No.:            3.[illegible]-23       Length:        3255
Score:                32.[illegible]         Matches:       32
Percent Similarity:   100.00%                Conservative:  0
Best Local Similarity: 100.00%               Mismatches:    0
Query Match:          3.27%                  Indels:        0
DB:                   10                     Gaps:          0

US-09-980-043-2 (1-978) x US-09-938-842A-559 (1-3255)

Qy        256 PheValSerThrValAspProLeuLysGluProProLeuIleThrAlaAsnThrValLeu 275
              ||||||||||||||[illegible]||||||||||||||||||||||||||||||||||||||
Db       1075 TTTGTTAGTACAGTGGATCCGTTGAAAGAGCCACCCTTGATTACAGCAAACACAGTTCTT 1134

Qy        276 SerIleLeuAlaValAspTyrProValAspLysVal 287
              ||||||||||||||||||||||||||||||||||||
```

```
Db       1135 TCCATTCTAGCAGTTGATTATCCTGTGGATAAGGTT 1170

RESULT 11
US-09-938-842A-2530
; Sequence 2530, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
;  APPLICANT: Harper, Jeff
;  APPLICANT:  Kreps, Joel
;  APPLICANT:  Wang, Xun
;  APPLICANT:  Zhu, Tong
;  TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
;  TITLE OF INVENTION:  SAME, AND METHODS OF USE
;  FILE REFERENCE: SCRIP1300-3
;  CURRENT APPLICATION NUMBER: US/09/938,842A
;  CURRENT FILING DATE:  2001-08-24
;  PRIOR APPLICATION NUMBER: US 60/227,866
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: US 60/264,647
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/300,111
;  PRIOR FILING DATE: 2001-06-22
;  NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 2530
;   LENGTH: 3255
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
US-09-938-842A-2530

Alignment Scores:
Pred. No.:                3.15e-23        Length:         3255
Score:                    32.00           Matches:        32
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              3.27%           Indels:         0
DB:                       10              Gaps:           0

US-09-980-043-2 (1-978) x US-09-938-842A-2530 (1-3255)

Qy        256 PheValSerThrValAspProLeuLysGluProProLeuIleThrAlaAsnThrValLeu 275
              ||| |||||||||||||||||||||| |||| |||| |||| |||| |||| ||||
Db       1092 TTTGTTAGTACAGTGGATCCACTGAAAGAGCCTCCGCTTATTACTGCAAATACTGTCTTG 1131

Qy        276 SerIleLeuAlaValAspTyrProValAspLysVal 287
              ||| |||| |||||||||| |||| |||| ||||
Db       1132 TCCATTCTTGCTGTTGATTATCCTGTGGATAAGGTT 1167

RESULT 12
US-10-229-193-7
; Sequence 7, Application US/10229193
; Publication No. US20030126643A1
; GENERAL INFORMATION:
;  APPLICANT: Arioli, Antonio Williamson, Richard E.
;  APPLICANT:  Betzner, Andreas S.
;  APPLICANT:  Peng, Liangcai
;  TITLE OF INVENTION: Manipulation of Cellulose and/or Beta-1.4-Glucan
;  FILE REFERENCE: 96-98A
;  CURRENT APPLICATION NUMBER: US/10/229,193
;  CURRENT FILING DATE:  2002-08-26
;  PRIOR APPLICATION NUMBER: 09/221,013
;  PRIOR FILING DATE: 1998-12-23
;  PRIOR APPLICATION NUMBER: PCT/AU97/00402
;  PRIOR FILING DATE: 1997-06-24
;  PRIOR APPLICATION NUMBER: AU PO0699
;  PRIOR FILING DATE: 1996-06-27
;  NUMBER OF SEQ ID NOS: 36
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
;   LENGTH: 3828
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (239)..(3490)
US-10-229-193-7

Alignment Scores:
Pred. No.:                3.67e-23        Length:         3828
Score:                    32.00           Matches:        32
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              3.27%           Indels:         0
DB:                       14              Gaps:           0

US-09-980-043-2 (1-978) x US-10-229-193-7 (1-3828)

Qy        256 PheValSerThrValAspProLeuLysGluProProLeuIleThrAlaAsnThrValLeu 275
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1313 TTTGTTAGTACAGTGGATCCGTTGAAAGAGCCCCCCTTGATTACAGCAAACACAGTTCTT 1372

Qy        276 SerIleLeuAlaValAspTyrProValAspLysVal 287
              ||||||||||||||||||||||||||||||||
Db       1373 TCCATTCTAGCAGTTGATTATCCTGTGGATAAGGTT 1408

RESULT 13
US-10-229-193-4
; Sequence 4, Application US/10229193
; Publication No. US20030126643A1
; GENERAL INFORMATION:
;  APPLICANT: Arioli, Antonio Williamson, Richard E.
;  APPLICANT:  Betzner, Andreas S.
;  APPLICANT:  Peng, Liangcai
;  TITLE OF INVENTION: Manipulation of Cellulose and/or Beta-1.4-Glucan
;  FILE REFERENCE: 96-98A
;  CURRENT APPLICATION NUMBER: US/10/229,193
;  CURRENT FILING DATE:  2002-08-26
;  PRIOR APPLICATION NUMBER: 09/221,013
;  PRIOR FILING DATE: 1998-12-23
;  PRIOR APPLICATION NUMBER: PCT/AU97/00402
;  PRIOR FILING DATE: 1997-06-24
;  PRIOR APPLICATION NUMBER: AU PO0699
;  PRIOR FILING DATE: 1996-06-27
;  NUMBER OF SEQ ID NOS: 36
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
;   LENGTH: 5009
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(5009)
;   OTHER INFORMATION: N is A, T, C or G.
US-10-229-193-4

Alignment Scores:
Pred. No.:                4.75e-23        Length:         5009
Score:                    32.00           Matches:        32
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              3.27%           Indels:         0
DB:                       14              Gaps:           0

US-09-980-043-2 (1-978) x US-10-229-193-4 (1-5009)

Qy        256 PheValSerThrValAspProLeuLysGluProProLeuIleThrAlaAsnThrValLeu 275
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2822 TTTGTTAGTACAGTGGATCCGTTGAAAGAGCCACCCTTGATTACAGCAAACACAGTTCTT 2881

Qy        276 SerIleLeuAlaValAspTyrProValAspLysVal 287
              ||||||||||||||||||||||||||||||||
Db       2882 TCCATTCTAGCAGTTGATTATCCTGTGGATAAGGTT 2917

RESULT 14
US-09-980-043-5
; Sequence 5, Application US/09980237
; Patent No. US20020120124A1
```

GENERAL INFORMATION:
APPLICANT: Allen, Stephen
TITLE OF INVENTION: Plant Cellulose Synthases
FILE REFERENCE: BB1170 US CIP
CURRENT APPLICATION NUMBER: US/09/900,237
CURRENT FILING DATE: 2001-07-06
PRIOR APPLICATION NUMBER: 60/092,844
PRIOR FILING DATE: 1998-07-14
PRIOR APPLICATION NUMBER: PCT/US99/15871
PRIOR FILING DATE: 1999-07-13
PRIOR APPLICATION NUMBER: 09/720383
PRIOR FILING DATE: 2000-12-21
NUMBER OF SEQ ID NOS: 33
SOFTWARE: Microsoft Office 97
SEQ ID NO 5
LENGTH: 1189
TYPE: DNA
ORGANISM: *Zea mays*
3-09-900-237-5

ignment Scores:

ed. No.:	1.59e-21	Length:	1189
core:	30.00	Matches:	30
ercent Similarity:	100.00%	Conservative:	0
est Local Similarity:	100.00%	Mismatches:	0
ery Match:	3.07%	Indels:	0
B:	10	Gaps:	0

3-09-980-043-2 (1-978) x US-09-900-237-5 (1-1189)

```
y        698 ArgProAlaPheLysGlySerAlaProIleAsnLeuSerAspArgLeuHisGlnValLeu 717
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
b        119 CGGCCGGCGTTCAAGGGGTCGGCGCCCATCAACTTGTCTGATCGTCTCCACCAGGTGCTG 178

y        718 ArgTrpAlaLeuGlySerValGluIlePhe 727
             |||||||||||||||||||||||||||
b        179 CGCTGGGCGCTGGGGTCCGTGGAGATCTTC 208
```

RESULT 15
S-09-900-237-23
Sequence 23, Application US/09900237
Patent No. US20020120124A1
GENERAL INFORMATION:
APPLICANT: Allen, Stephen
TITLE OF INVENTION: Plant Cellulose Synthases
FILE REFERENCE: BB1170 US CIP
CURRENT APPLICATION NUMBER: US/09/900,237
CURRENT FILING DATE: 2001-07-06
PRIOR APPLICATION NUMBER: 60/092,844
PRIOR FILING DATE: 1998-07-14
PRIOR APPLICATION NUMBER: PCT/US99/15871
PRIOR FILING DATE: 1999-07-13
PRIOR APPLICATION NUMBER: 09/720383
PRIOR FILING DATE: 2000-12-21
NUMBER OF SEQ ID NOS: 33
SOFTWARE: Microsoft Office 97
SEQ ID NO 23
LENGTH: 2663
TYPE: DNA
ORGANISM: Picramnia pentandra
US-09-900-237-23

Alignment Scores:

Pred. No.:	3.43e-21	Length:	2663
Score:	30.00	Matches:	30
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	3.07%	Indels:	0
DB:	10	Gaps:	0

US-09-980-043-2 (1-978) x US-09-900-237-23 (1-2663)

```
Qy       698 ArgProAlaPheLysGlySerAlaProIleAsnLeuSerAspArgLeuHisGlnValLeu 717

Db      1386 CGCCCGCATTTAA[illegible]TTCAGATCGTCTACATCAGGTTCTT 1443

Qy       718 ArgTrpAlaLeuGlySerValGluIlePhe 727

Db      1446 CGGTGGGCCCTG[illegible]TGGAGATCTTC 1476
```

Search completed: November [illegible], 2003, [illegible]:42
Job time : 772.215 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame plus p2n model

Run on: October 31, 2003, 2:41:14 ; Search time 156.064 Seconds
(without alignments)
2766.681 Million cell updates/sec

Title: US-09-980-043-2
Perfect score: 978
Sequence: 1 MMESGAPICHTCGEQVGSDA.....IFVNEVLVELVAETCISIDC 978

Scoring table: OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 669978 seqs, 220691566 residues

Word size: 1

Total number of hits satisfying chosen parameters: 1135299

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:
MODEL=frame+ p2n model -DEV=xlh
Q=/cgn2_1/USPTO_spool/US09980043/runat_29102003_163931_22725/app_query.fasta_1.1358
-DB=Issued_Patents_NA -QFMT=fastap SUFFIX oligo.rni MINMATCH=0.1 -LOOPCL=0
LOOPEXT=0 -UNITS=bits -START=1 END=-1 MATRIX=oligo "@ANS=human40.cdi
LIST 45 DOCALIGN=200 THR SCORE=quality THR MIN=1 ALIGN=15 -MODE=LOCAL
OUTFMT=pto -NORM=ext HEAPSIZE=500 MINLEN 0 -MAXLEN=2000000000
USER US09980043 aCGN_1_1_76_runat_29102003_163931_22725 NCPU=6 ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 LONGLOG
DEV TIMECUT=120 WARN_TIMECUT=30 -THREADS=1 XGAPOP 60 XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP 60 YGAPEXT=60 -DELOP=6 DELEXT=7

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No	Score	Query Match	Length	DB	ID	Description
1	61	6.2	3328	3	US-08-960-048-1	Sequence 1, Appli
2	61	6.2	3328	4	US-09-836-586-1	Sequence 1, Appli
3	40	4.1	1741	4	US-09-221-013A-13	Sequence 13, Appl
4	35	3.6	3614	4	US-09-221-013A-9	Sequence 9, Appli
5	32	3.3	3828	4	US-09-221-013A-7	Sequence 7, Appli
6	32	3.3	5009	4	US-09-221-013A-4	Sequence 4, Appli
7	23	2.4	3603	4	US-09-221-013A-5	Sequence 5, Appli
8	23	2.4	3673	4	US-09-221-013A-11	Sequence 11, Appl
9	23	2.4	4612	3	US-08-960-048-2	Sequence 2, Appli
10	23	2.4	4612	4	US-09-836-586-2	Sequence 2, Appli
11	23	2.4	8411	4	US-09-221-013A-3	Sequence 3, Appli
12	21	2.1	2248	4	US-09-221-013A-1	Sequence 1, Appli
13	17	1.7	283	4	US-09-313-294A-1981	Sequence 1981, Ap
14	15	1.5	284	4	US-09-313-294A-3993	Sequence 3993, Ap
15	13	1.3	268	4	US-09-313-294A-849	Sequence 849, App
16	13	1.3	278	4	US-09-313-294A-4015	Sequence 4015, Ap
17	13	1.3	291	4	US-09-313-294A-5519	Sequence 5519, Ap
18	12	1.2	480	4	US-09-276-599-8	Sequence 8, Appli
19	12	1.2	591	4	US-09-276-599-7	Sequence 7, Appli
20	11	1.1	281	4	US-09-313-294A-6485	Sequence 6485, Ap
21	11	1.1	286	4	US-09-313-294A-4696	Sequence 4696, Ap
22	11	1.1	300	4	US-09-313-294A-4753	Sequence 4753, Ap
23	9	0.9	159	4	US-09-313-294A-1693	Sequence 1693, Ap
c 24	9	0.9	3820	3	US-08-990-140-3	Sequence 3, Appli
c 25	9	0.9	3820	4	US-09-546-238-3	Sequence 3, Appli
c 26	9	0.9	3829	2	US-08-631-097-8	Sequence 8, Appli
c 27	9	0.9	3829	3	US-08-810-712-6	Sequence 6, Appli
28	8	0.8	250	4	US-09-313-294A-582	Sequence 582, App
c 29	8	0.8	886	3	US-08-924-747-1	Sequence 1, Appli
c 30	8	0.8	886	3	US-09-247-373B-1	Sequence 1, Appli
c 31	8	0.8	886	3	US-09-296-715-1	Sequence 1, Appli
c 32	8	0.8	1986	4	US-09-107-532A-23	Sequence 23, Appl
33	8	0.8	2187	4	US-09-134-001C-1429	Sequence 1429, Ap
c 34	8	0.8	2229	4	US-09-107-532A-1342	Sequence 1342, Ap
c 35	8	0.8	2809	3	US-08-448-194-5	Sequence 5, Appli
c 36	8	0.8	2809	4	US-08-867-921-5	Sequence 5, Appli
c 37	8	0.8	3286	2	US-08-363-124A-1	Sequence 1, Appli
c 38	8	0.8	3471	4	US-09-221-017B-41	Sequence 41, Appl
39	8	0.8	11272	4	US-09-341-461-1	Sequence 1, Appli
c 40	8	0.8	26664	4	US-09-564-805-28	Sequence 28, Appl
c 41	8	0.8	129908	4	US-09-585-858-1	Sequence 1, Appli
42	7	0.7	22	4	US-09-221-013A-27	Sequence 27, Appl
c 43	7	0.7	23	4	US-09-221-013A-32	Sequence 32, Appl
44	7	0.7	24	4	US-09-221-013A-26	Sequence 26, Appl
45	7	0.7	30	3	US-09-178-869-11	Sequence 11, Appl

ALIGNMENTS

RESULT 1
US-08-960-048-1
; Sequence 1, Application US/08960048C
; Patent No. 6271443
; GENERAL INFORMATION:
; APPLICANT: Stalker, D. et al.
; TITLE OF INVENTION: Plant Cellulose Synthase and Promoter
; TITLE OF INVENTION: Sequences
; FILE REFERENCE: 15621/01/US
; CURRENT APPLICATION NUMBER: US/08/960,048C
; CURRENT FILING DATE: 1997-10-29
; PRIOR APPLICATION NUMBER: 60/029,987
; PRIOR FILING DATE: 1996-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 3328
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Oligonucleotide
US-08-960-048-1

Alignment Scores:

Pred. No.:	2.52e-54	Length:	3328
Score:	61.00	Matches:	61
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	6.24%	Indels:	0
DB:	3	Gaps:	0

US-09-980-043-2 (1-978) x US-08-960-048-1 (1-3328)

```
Qy        799 SerIleIleValThrAlaValLeuGluLeuArgTrpSerGlyValSerIleGluAspLeu 818
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2558 TCCATTATCGTGACTGCTGTTCTCGAGCTCCGATGGAGTGGTGTCAGCATTGAGGACTTA 2617
```

```
        819 TrpArgAsnGluGlnPheTrpValIleGlyGlyValSerAlaHisLeuPheAlaValPhe 838
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
>      2618 TGGCGTAACGAGCAGTTTTGGGTCATCGGTGGCGTTTCAGCCCATCTCTTTGCCGTCTTC 2677

/       839 GlnGlyPheLeuLysMetLeuAlaGlyIleAspThrAsnPheThrValThrAlaLysAla 858
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
>      2678 CAAGGTTTCCTTAAGATGCTTGCGGGCATTGACACCAACTTTACTGTCACTGCCAAAGCA 2737

/       859 Ala 859
            |||
>      2738 GCT 2740

ESULT 2
S-09-838-586-1
 Sequence 1, Application US/09838586
 Patent No. 6576818
 GENERAL INFORMATION:
  APPLICANT: Stalker, D. et al.
  TITLE OF INVENTION: Plant Cellulose Synthase and Promoter
  TITLE OF INVENTION:  Sequences
  FILE REFERENCE: 15621/02/US
  CURRENT APPLICATION NUMBER: US/09/838,586
  CURRENT FILING DATE:  2001-04-18
  PRIOR APPLICATION NUMBER: 60/029,987
  PRIOR FILING DATE: 1996-10-29
  PRIOR APPLICATION NUMBER: 08/960,048
  PRIOR FILING DATE: 1997-10-29
  NUMBER OF SEQ ID NOS: 12
  SOFTWARE: FastSEQ for Windows Version 4.0
 SEQ ID NO 1
   LENGTH: 3328
   TYPE: DNA
   ORGANISM: Artificial Sequence
   FEATURE:
   OTHER INFORMATION: Synthetic Oligonucleotide
S-09-838-586-1

lignment Scores:
red. No.:               2.52e-54        Length:         3328
core:                   61.00           Matches:        61
ercent Similarity:      100.00%         Conservative:   0
est Local Similarity:   100.00%         Mismatches:     0
uery Match:             6.24%           Indels:         0
B:                      4               Gaps:           0

S-09-980-043-2 (1-978) x US-09-838-586-1 (1-3328)

y       799 SerIleIleValThrAlaValLeuGluLeuArgTrpSerGlyValSerIleGluAspLeu 818
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
b      2558 TCCATTATCGTGACTGCTGTTCTCGAGCTCCGATGGAGTGGTGTCAGCATTGAGGACTTA 2617

y       819 TrpArgAsnGluGlnPheTrpValIleGlyGlyValSerAlaHisLeuPheAlaValPhe 838
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
b      2618 TGGCGTAACGAGCAGTTTTGGGTCATCGGTGGCGTTTCAGCCCATCTCTTTGCCGTCTTC 2677

y       839 GlnGlyPheLeuLysMetLeuAlaGlyIleAspThrAsnPheThrValThrAlaLysAla 858
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
b      2678 CAAGGTTTCCTTAAGATGCTTGCGGGCATTGACACCAACTTTACTGTCACTGCCAAAGCA 2737

y       859 Ala 859
            |||
b      2738 GCT 2740

ESULT 3
S-09-221-013A-13
 Sequence 13, Application US/09221013A
 Patent No. 6495740
 GENERAL INFORMATION:
  APPLICANT:  Arioli, Antonio
  APPLICANT:  Williamson, Richard E.
  APPLICANT:  Betzner, Andreas S.
  APPLICANT:  Peng, Liangcai
;  TITLE OF INVENTION:  Manipulation of Cellulose and/or Beta 1,4-Glucan
;  FILE REFERENCE:  96-98
;  CURRENT APPLICATION NUMBER:  US/09/221,013A
;  CURRENT FILING DATE:  1998-12-23
;  PRIOR APPLICATION NUMBER:  PCT/AU97/00402
;  PRIOR FILING DATE:  1997-06-24
;  PRIOR APPLICATION NUMBER:  AU PO0699
;  PRIOR FILING DATE:  1996-06-27
;  NUMBER OF SEQ ID NOS:  13
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 13
;   LENGTH: 1741
;   TYPE: DNA
;   ORGANISM: Oryza sativa
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (101)..(1741)
US-09-221-013A-13

Alignment Scores:
Pred. No.:              2.7e-54         Length:         1741
Score:                  40              Matches:        40
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            4.09%           Indels:         0
DB:                     4               Gaps:           0

US-09-980-043-2 (1-978) x US-09-221-013A-13 (101-1741)

Qy        256 PheValSerThrValAspProIleLysGluProProLeuIleThrAlaAsnThrValLeu 275
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1160 TTTGTCAGTACGGTGGATCCAATAAAGGAACCTCCTCTGATCACAGCAAACACTGTTTTG 1219

Qy        276 SerIleLeuAlaValAspTyrProValAspLysValSerCysTyrValSerAspAspGly 295
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1220 TCCATTCTGGCTGTGGATTACCCTGTTGACAAAGTGTCATGCTATGTTTCTGACGATGGT 1279

RESULT 4
US-09-221-013A-9
; Sequence 9, Application US/09221013A
; Patent No. 6495740
; GENERAL INFORMATION:
;  APPLICANT:  Arioli, Antonio
;  APPLICANT:  Williamson, Richard E.
;  APPLICANT:  Betzner, Andreas S.
;  APPLICANT:  Peng, Liangcai
;  TITLE OF INVENTION:  Manipulation of Cellulose and/or Beta-1,4 Glucan
;  FILE REFERENCE:  96-98
;  CURRENT APPLICATION NUMBER:  US/09/221,013A
;  CURRENT FILING DATE:  1998-12-23
;  PRIOR APPLICATION NUMBER:  PCT/AU97/00402
;  PRIOR FILING DATE:  1997-06-24
;  PRIOR APPLICATION NUMBER:  AU PO0699
;  PRIOR FILING DATE:  1996-06-27
;  NUMBER OF SEQ ID NOS:  13
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 9
;   LENGTH: 3614
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (217)..(3411)
US-09-221-013A-9

Alignment Scores:
Pred. No.:              9.01e-27        Length:         3614
Score:                  35.00           Matches:        35
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            3.58%           Indels:         0
DB:                     4               Gaps:           0
```

US-09-980-043-2 (1-978) x US-09-221-013A-9 (1-3614)

Qy 270 ThrAlaAsnThrValLeuSerIleLeuAlaValAspTyrProValAspLysValSerCys 289
Db 1279 ACAGCCAACACAGTGCTCTCTATTCTGGCTGTTGACTACCCAGTTGACAAGGTGTCCTGT 1338

Qy 290 TyrValSerAspAspGlyAlaAlaMetLeuSerPheGluSerLeu 304
Db 1339 TATGTTTCTGATGATGGTGCTGCTATGTTATCATTTGAATCACTT 1383

RESULT 5
US-09-221-013A-7
; Sequence 7, Application US/09221013A
; Patent No. 6495740
; GENERAL INFORMATION:
; APPLICANT: Arioli, Antonio
; APPLICANT: Williamson, Richard E
; APPLICANT: Betzner, Andreas S
; APPLICANT: Peng, Liangcai
; TITLE OF INVENTION: Manipulation of cellulose and/or Beta-1,4-glucan
; FILE REFERENCE: 96-98
; CURRENT APPLICATION NUMBER: US/09/221,013A
; CURRENT FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: PCT/AU97/00402
; PRIOR FILING DATE: 1997-06-24
; PRIOR APPLICATION NUMBER: AU PO0699
; PRIOR FILING DATE: 1996-06-27
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 3828
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (239)..(3490)
US-09-221-013A-7

Alignment Scores:
Pred. No.: 1.43e-23 Length: 3828
Score: 32.00 Matches: 32
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.27% Indels: 0
DB: 4 Gaps: 0

US-09-980-043-2 (1-978) x US-09-221-013A-7 (1-3828)

Qy 256 PheValSerThrValAspProLeuLysGluProProLeuIleThrAlaAsnThrValLeu 275
Db 1313 TTTGTTAGTACAGTGGATCCGTTGAAAGAGCCACCCTTGATTACAGCAAACACAGTTCTT 1372

Qy 276 SerIleLeuAlaValAspTyrProValAspLysVal 287
Db 1373 TCCATTCTAGCAGTTGATTATCCTGTGGATAAGGTT 1408

RESULT 6
US-09-221-013A-4
; Sequence 4, Application US/09221013A
; Patent No. 6495740
; GENERAL INFORMATION:
; APPLICANT: Arioli, Antonio
; APPLICANT: Williamson, Richard E.
; APPLICANT: Betzner, Andreas S.
; APPLICANT: Peng, Liangcai
; TITLE OF INVENTION: Manipulation of cellulose and/or Beta-1,4-glucan
; FILE REFERENCE: 96-98
; CURRENT APPLICATION NUMBER: US/09/221,013A
; CURRENT FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: PCT/AU97/00402
; PRIOR FILING DATE: 1997-06-24
; PRIOR APPLICATION NUMBER: AU PO0699
; PRIOR FILING DATE: 1996-06-27
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 5009
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(5009)
; OTHER INFORMATION: N is A, T, G or C.
US-09-221-013A-4

Alignment Scores:
Pred. No.: 1.85e-23 Length: 5009
Score: 32.00 Matches: 32
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.27% Indels: 0
DB: 4 Gaps: 0

US-09-980-043-2 (1-978) x US-09-221-013A-4 (1-5009)

Qy 256 PheValSerThrValAspProLeuLysGluProProLeuIleThrAlaAsnThrValLeu 275
Db 2822 TTTGTTAGTACAGTGGATCCGTTGAAAGAGCCACCCTTGATTACAGCAAACACAGTTCTT 2881

Qy 276 SerIleLeuAlaValAspTyrProValAspLysVal 287
Db 2882 TCCATTCTAGCAGTTGATTATCCTGTGGATAAGGTT 2917

RESULT 7
US-09-221-013A-5
; Sequence 5, Application US/09221013A
; Patent No. 6495740
; GENERAL INFORMATION:
; APPLICANT: Arioli, Antonio
; APPLICANT: Williamson, Richard E.
; APPLICANT: Betzner, Andreas S.
; APPLICANT: Peng, Liangcai
; TITLE OF INVENTION: Manipulation of cellulose and/or Beta-1,4-glucan
; FILE REFERENCE: 96-98
; CURRENT APPLICATION NUMBER: US/09/221,013A
; CURRENT FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: PCT/AU97/00402
; PRIOR FILING DATE: 1997-06-24
; PRIOR APPLICATION NUMBER: AU PO0699
; PRIOR FILING DATE: 1996-06-27
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 3603
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(3243)
US-09-221-013A-5

Alignment Scores:
Pred. No.: 4.51e-14 Length: 3603
Score: 23.00 Matches: 23
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.35% Indels: 0
DB: 4 Gaps: 0

US-09-980-043-2 (1-978) x US-09-221-013A-5 (1-3603)

Qy 376 ProGluGluGlyTrpThrMetGlnAspGlyThrProTrpProGlyAsnAsnThrArgAsp 395
Db 1429 CCTGAAGAAGGCTGGACAATGCAGGATGGTACTCCCTGGCCTGGTAACAACACTAGAGAT 1488

Qy 396 HisProGly 398

```
                ||||||||||
Db       1489 CATCCTGGA 1497

ESULT 8
S-09-221-013A-11
 Sequence 11, Application US/09221013A
 Patent No. 6495740
 GENERAL INFORMATION:
  APPLICANT:  Arioli, Antonio
  APPLICANT:  Williamson, Richard E.
  APPLICANT:  Betzner, Andreas S.
  APPLICANT:  Peng, Liangcai
  TITLE OF INVENTION:  Manipulation of cellulose and/or Beta 1,4 glucan
  FILE REFERENCE:  96-98
  CURRENT APPLICATION NUMBER: US/09/221,013A
  CURRENT FILING DATE:  1998-12-23
  PRIOR APPLICATION NUMBER:  PCT/AU97/00402
  PRIOR FILING DATE:  1997-06-24
  PRIOR APPLICATION NUMBER:  AU PO0699
  PRIOR FILING DATE:  1996-06-27
  NUMBER OF SEQ ID NOS:  37
  SOFTWARE:  PatentIn Ver. 2.0
 SEQ ID NO 11
    LENGTH: 3673
    TYPE: DNA
    ORGANISM: Arabidopsis thaliana
    FEATURE:
    NAME/KEY: CDS
    LOCATION: (71)..(3313)
S-09-221-013A-11

lignment Scores:
red. No.:                4.59e-14        Length:          3673
core:                    23.00           Matches:         23
ercent Similarity:       100.00%         Conservative:    0
est Local Similarity:    100.00%         Mismatches:      0
uery Match:              2.35%           Indels:          0
B:                       4               Gaps:            0

S-09-980-043 2 (1-978) x US-09-221-013A-11 (1-3673)

Qy         376 ProGluGluGlyTrpThrMetGlnAspGlyThrProTrpProGlyAsnAsnThrArgAsp 395
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1499 CCTGAAGAAGGCTGGACAATGCAGGATGGTACTCCCTGGCCTGGTAACAACACTAGAGAT 1558

Qy         396 HisProGly 398
               |||||||||
Db        1559 CATCCTGGA 1567

ESULT 9
S-08-960-048 2
 Sequence 2, Application US/08960048C
 Patent No. 6271443
 GENERAL INFORMATION:
  APPLICANT:  Stalker, D. et al.
  TITLE OF INVENTION:  Plant Cellulose Synthase and Promoter
  TITLE OF INVENTION:  Sequences
  FILE REFERENCE:  15621/01/US
  CURRENT APPLICATION NUMBER: US/08/960,048C
  CURRENT FILING DATE:  1997-10-29
  PRIOR APPLICATION NUMBER:  60/029,987
  PRIOR FILING DATE:  1996-10-29
  NUMBER OF SEQ ID NOS:  12
  SOFTWARE:  FastSEQ for Windows Version 4.0
 SEQ ID NO 2
    LENGTH: 4612
    TYPE: DNA
    ORGANISM: Artificial Sequence
    FEATURE:
    OTHER INFORMATION: Synthetic Oligonucleotide
S-08-960-048-2

lignment Scores:

Pred. No.:                5.7e-14         Length:          4612
Score:                    23.00           Matches:         23
Percent Similarity:       100.00 %        Conservative:    0
Best Local Similarity:    100.00 %        Mismatches:      0
Query Match:              2.35%           Indels:          0
DB:                       4               Gaps:            0

US-09-980-043-2 (1-978) x US-09-960-048-2 (1-4612)

Qy         699 ProAlaPheLysGlySerAlaProIleAsnLeuSerAspArgLeuHisGlnValLeuArg 718
               |||| ||| ||| ||| ||| ||| ||| |||| |||| |||| |||| |||| |||| |
Db        2523 CCGGCATTCAAAGGTTCGGCTCCAATCAATCTTTCGGATCGGTTGCACCAAGTTTTCAGA 2582

Qy         719 TrpAlaLeu 727
               |||:|||
Db        2583 TGGGCACTC 2591

RESULT 10
US-09-838-586-2
; Sequence 2, Application US/09838586
; Patent No. 6576818
; GENERAL INFORMATION:
;  APPLICANT: Stalker, D. et al.
;  TITLE OF INVENTION: Plant Cellulose Synthase and Promoter
;  TITLE OF INVENTION:  Sequences
;  FILE REFERENCE: 15621/02/US
;  CURRENT APPLICATION NUMBER: US/09/838,586
;  CURRENT FILING DATE:  2001-04-18
;  PRIOR APPLICATION NUMBER: 60/029,987
;  PRIOR FILING DATE: 1996-10-29
;  PRIOR APPLICATION NUMBER: 08/960,048
;  PRIOR FILING DATE: 1997-10-29
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 4612
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic Oligonucleotide
US-09-838-586-2

Alignment Scores:
Pred. No.:                5.7e-14         Length:          4612
Score:                    23.00           Matches:         23
Percent Similarity:       100.00 %        Conservative:    0
Best Local Similarity:    100.00 %        Mismatches:      0
Query Match:              2.35%           Indels:          0
DB:                       4               Gaps:            0

US-09-980-043-2 (1-978) x US-09-838-586-2 (1-4612)

Qy         699 ProAlaPheLysGlySerAlaProIleAsnLeuSerAspArgLeuHisGlnValLeuArg 718
               |||| ||| ||| ||| ||| ||| ||| |||| |||| |||| |||| |||| |||| |
Db        2523 CCGGCATTCAAAGGTTCGGCTCCAATCAATCTTTCGGATCGGTTGCACCAAGTTTTCAGA 2582

Qy         719 TrpAlaLeu 727
               |: |||| |
Db        2583 TGGGCACTC 2591

RESULT 11
US-09-221-013A-3
; Sequence 3, Application US/09221013A
; Patent No. 6495740
; GENERAL INFORMATION:
;  APPLICANT:  Arioli, Antonio
;  APPLICANT:  Williamson, Richard E.
;  APPLICANT:  Betzner, Andreas S.
;  APPLICANT:  Peng, Liangcai
;  TITLE OF INVENTION:  Manipulation of cellulose and/or Beta 1,4-glucan
;  FILE REFERENCE:  96-98
;  CURRENT APPLICATION NUMBER: US/09/221,013A
;  CURRENT FILING DATE:  1998-12-23
```

```
;  PRIOR APPLICATION NUMBER:  PCT/AU97/00402
;  PRIOR FILING DATE:  1997 06 24
;  PRIOR APPLICATION NUMBER:  AU PO0699
;  PRIOR FILING DATE:  1996-06-27
;  NUMBER OF SEQ ID NOS:  37
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 8411
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
US-09-221-013A 3

Alignment Scores:
Pred. No.:                1.03e-13        Length:          8411
Score:                    23.00           Matches:            2
Percent Similarity:       100.00%         Conservative:       0
Best Local Similarity:    100.00%         Mismatches:         0
Query Match:              2.35%           Indels:             0
DB:                       4               Gaps:               0

US-09-980-043-2 (1-978) x US-09-221-013A 3 (1-8411)

Qy         376 ProGluGluGlyTrpThrMetGlnAspGlyThrProTrpProGlyAsnAsnThrArgAsp 395
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        5267 CCTGAAGAAGGCTGGACAATGCAGGATGGTACTCCCTGGCCTGGTAACAACACTAGAGAT 5326

Qy         396 HisProGly 398
               |||||||
Db        5327 CATCCTGGA 5335

RESULT 12
US-09-221-013A-1
; Sequence 1, Application US/09221013A
; Patent No. 6495740
; GENERAL INFORMATION:
;  APPLICANT:  Arioli, Antonio
;  APPLICANT:  Williamson, Richard E
;  APPLICANT:  Betzner, Andreas S.
;  APPLICANT:  Peng, Liangcai
;  TITLE OF INVENTION:  Manipulation of Cellulose and/or Beta 1,4 glucan
;  FILE REFERENCE:  96-98
;  CURRENT APPLICATION NUMBER: US/09/221,013A
;  CURRENT FILING DATE:  1998-12-23
;  PRIOR APPLICATION NUMBER:  PCT/AU97/00402
;  PRIOR FILING DATE:  1997-06-24
;  PRIOR APPLICATION NUMBER:  AU PO0699
;  PRIOR FILING DATE:  1996-06-27
;  NUMBER OF SEQ ID NOS:  37
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 2248
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(1887)
US-09-221-013A 1

Alignment Scores:
Pred. No.:                3.72e-12        Length:          2248
Score:                    21.00           Matches:            3
Percent Similarity:       100.00%         Conservative:       0
Best Local Similarity:    100.00%         Mismatches:         0
Query Match:              2.15%           Indels:             0
DB:                       4               Gaps:               0

US-09-980-043-2 (1-978) x US-09-221-013A 1 (1-2248)

Qy         378 GluGlyTrpThrMetGlnAspGlyThrProTrpProGlyAsnAsnThrArgAspHisPro 397
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          79 GAAGGCTGGACAATGCAGGATGGTACTCCCTGGCCTGGTAACAACACTAGAGATCATCCT 138

Qy         398 Gly 398
               |||
Db         139 GGA 141

RESULT 13
US-09-313-294A-1981
; Sequence 1981, Application US/09313294A
; Patent No. 6476212
; GENERAL INFORMATION:
;  APPLICANT:  Lalgudi, Raghunath V.
;  APPLICANT:  Ito, Laura Y.
;  APPLICANT:  Sherman, Bradley K.
;  TITLE OF INVENTION:  POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
;  FILE REFERENCE:  PL-0017 US
;  CURRENT APPLICATION NUMBER: US/09/313,294A
;  CURRENT FILING DATE:  1999-05-14
;  NUMBER OF SEQ ID NOS:  7600
;  SOFTWARE:  PERL Program
; SEQ ID NO 1981
;   LENGTH: 283
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. 6476212 700551845H1
;   NAME/KEY: unsure
;   LOCATION: 55, 189
;   OTHER INFORMATION: a, t, c, g, or other
US-09-313-294A-1981

Alignment Scores:
Pred. No.:                8.46e-09        Length:          283
Score:                    17.00           Matches:          17
Percent Similarity:       100.00%         Conservative:      0
Best Local Similarity:    100.00%         Mismatches:        0
Query Match:              1.74%           Indels:            0
DB:                       4               Gaps:              0

US-09-980-043-2 (1-978) x US-09-313-294A-1981 (1-283)

Qy         715 GlnValLeuArgTrpAlaLeuGlySerValGluIlePhePheSerArgHis 731
               ||||||||||||||||||||||||||||||||||||||||||||||||||
Db           2 CAAGTTCTCCGGTGGGCTCTTGGTTCTGTCGAGATCTTCTTCAGTCGGCAC 52

RESULT 14
US 09-313-294A-3993
; Sequence 3993, Application US/09313294A
; Patent No. 6476212
; GENERAL INFORMATION:
;  APPLICANT:  Lalgudi, Raghunath V.
;  APPLICANT:  Ito, Laura Y.
;  APPLICANT:  Sherman, Bradley K.
;  TITLE OF INVENTION:  POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
;  FILE REFERENCE:  PL-0017 US
;  CURRENT APPLICATION NUMBER: US/09/313,294A
;  CURRENT FILING DATE:  1999-05-14
;  NUMBER OF SEQ ID NOS:  7600
;  SOFTWARE:  PERL Program
; SEQ ID NO 3993
;   LENGTH: 284
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. 6476212 700347494H1
;   NAME/KEY: unsure
;   LOCATION: 52, 189, 203, 209, 227, 240, 251, 254, 259, 264, 267, 278-279
;   OTHER INFORMATION: a, t, c, g, or other
US-09-313-294A-3993

Alignment Scores:
Pred. No.:                1.11e-06        Length:          284
Score:                    15.00           Matches:          15
Percent Similarity:       100.00%         Conservative:      0
```

```
est Local Similarity:  100.00%        Mismatches:    0
uery Match:            1.53%          Indels:        0
B:                     4              Gaps:          0

S-09-980-043-2 (1-978) x US-09-313-294A-3993 (1-284)

y        919 LeuHisLeuTyrProPheLeuLysGlyLeuMetGlyArgGlnAsn 933
             ||||||||||||||||||||||||||||||||||||||||||||||||
o         59 TTACACCTGTACCCGTTCCTCAAGGGTCTCATGGGTCGGCAGAAC 103

ESULT 15
S-09-313-294A-849
 Sequence 849, Application US/09313294A
 Patent No. 6476212
 GENERAL INFORMATION:
  APPLICANT:  Lalgudi, Raghunath V.
  APPLICANT:  Ito, Laura Y.
  APPLICANT:  Sherman, Bradley K.
  TITLE OF INVENTION:  POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
  FILE REFERENCE:  PL-0017 US
  CURRENT APPLICATION NUMBER: US/09/313,294A
  CURRENT FILING DATE:  1999-05-14
  NUMBER OF SEQ ID NOS:  7600
  SOFTWARE:  PERL Program
 SEQ ID NO 849
   LENGTH: 268
   TYPE: DNA
   ORGANISM: Zea mays
   FEATURE:
   NAME/KEY: misc_feature
   OTHER INFORMATION: Incyte ID No. 6476212 700549940H1
S-09-313 294A-849

lignment Scores:
red. No.:              0.000137       Length:        268
core:                  13.00          Matches:       13
ercent Similarity:     100.00%        Conservative:  0
est Local Similarity:  100.00%        Mismatches:    0
uery Match:            1.33%          Indels:        0
B:                     4              Gaps:          0

S-09-980-043-2 (1-978) x US-09-313-294A-849 (1-268)

y        819 TrpArgAsnGluGlnPheTrpValIleGlyGlyValSer 831
             ||||||||||||||||||||||||||||||||||||||||
b         74 TGGAGGAATGAGCAGTTCTGGGTCATTGGAGGTGTGTCC 112


earch completed: November  1, 2003, 04:48:03
ob time : 184.064 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 31, 2003, 17:45:37 ; Search time 719.785 Seconds
(without alignments)
3667.830 Million cell updates/sec

Title: US-09-980-043-2
Perfect score: 978
Sequence: 1 MMESGAPICHTCGEQVGHDA.....PFVNKVCGTLVAETCISIDC 978

Scoring table: OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Word size: 1

Total number of hits satisfying chosen parameters: 5103490

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:
-MODEL=frame_p2n.model -DEV=x11
Q=/cgn2_1/USPTO_spool/US09980043/runat_29102003_163930_22655/app_query.fasta_1.1358
-DB=N_Geneseq_19Jun03 -QFMT=fastap -SUFFIX=oligo.rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09980043 -CGN_1_1_458_@runat_29102003_163930_22655 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=-1 -XGAPOP=6 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : N_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqn-em:/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-em:/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-em:/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-em:/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-em:/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-em:/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-em:/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-em:/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-em:/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-em:/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-em:/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-em:/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-em:/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-em:/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-em:/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-em:/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-em:/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-em:/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-em:/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-em:/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-em:/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-em:/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-em:/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-em:/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseqn-em:/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	978	100.0	3232	22	AAC65448	Populus tremuloide
2	61	6.2	3207	19	AAV08372	Cellulose synthase
3	61	6.2	3328	19	AAV34432	Cotton cellulose s
4	44	4.5	1010	22	AAC65449	Populus tremuloide
5	43	4.4	950	21	AAC50201	Arabidopsis thalia
6	43	4.4	951	21	AAC33457	Arabidopsis thalia
7	43	4.4	3747	21	AAA67145	Eucalyptus grandis
8	40	4.1	719	21	AAA67084	Eucalyptus grandis
9	40	4.1	1741	19	AAV06569	Rice cellulose syn
10	40	4.1	3725	21	AAZ99500	DNA encoding a mai
11	40	4.1	3725	21	AAZ99506	DNA encoding a mai
12	40	4.1	3725	21	AAZ99521	DNA encoding a mai
13	39	4.0	1407	21	AAA67086	Eucalyptus grandis
14	39	4.0	2306	19	AAV34433	Cotton cellulose s
15	39	4.0	3311	19	AAV08373	Cellulose synthase
16	36	3.7	2890	21	AAZ58270	Soybean cellulose
17	35	3.6	3198	21	AAC49550	Arabidopsis thalia
18	35	3.6	3614	19	AAV06567	Arabidopsis cellul
19	32	3.3	3255	24	ABZ12754	Arabidopsis thalia
20	32	3.3	3255	24	ABZ14725	Arabidopsis thalia
21	32	3.3	3444	22	AAC83798	Arabidopsis thalia
22	32	3.3	3828	19	AAV06566	Arabidopsis cellul
23	32	3.3	5009	19	AAV06564	Arabidopsis cellul
24	30	3.1	590	21	AAA67136	Eucalyptus grandis
25	30	3.1	1189	21	AAZ58264	Corn cellulose syn
26	30	3.1	1860	21	AAA67081	Eucalyptus grandis
27	30	3.1	2055	21	AAA67111	Eucalyptus grandis
28	30	3.1	3704	21	AAZ99533	DNA encoding a mai
29	30	3.1	3746	21	AAZ99512	DNA encoding a mai
30	30	3.1	3746	21	AAZ99527	DNA encoding a mai
31	30	3.1	3773	21	AAZ99494	DNA encoding a mai
32	30	3.1	3799	24	AAS16458	Corn cDNA encoding
33	28	2.9	383	21	AAA67080	Pinus radiata cell
34	26	2.7	619	21	AAA67083	Pinus radiata cell
35	24	2.5	2830	24	AAS16455	Corn cDNA encoding
36	24	2.5	3517	21	AAZ58268	Soybean cellulose
37	24	2.5	7234	22	AAC83247	Cellulose synthase
38	23	2.4	3603	19	AAV06565	Arabidopsis cellul
39	23	2.4	3673	19	AAV06568	Arabidopsis cellul
40	23	2.4	3753	21	AAZ99515	DNA encoding a mai
41	23	2.4	3753	21	AAZ99530	DNA encoding a mai
42	23	2.4	3780	21	AAZ99497	DNA encoding a mai
43	23	2.4	8411	19	AAV06563	Arabidopsis cellul
44	22	2.2	1401	21	AAC44966	Arabidopsis thalia
45	22	2.2	3786	21	AAZ58265	Corn cellulose syn

ALIGNMENTS

RESULT 1
AAC65448
ID AAC65448 standard; cDNA; 3232 BP.
XX
AC AAC65448;
XX
DT 07-MAR-2001 (first entry)
XX
DE Populus tremuloides CelA cDNA.
XX
KW Aspen; CelA; cellulose synthase; arabidopsis; wood;
KW mechanical stress response element; MSRE; cellulose biosynthesis; ss.
XX
OS Populus tremuloides.
XX
PN WO200071670-A2.
XX

30-NOV-2000.

18-MAY-2000; 2000WO-US13637.

21-MAY-1999; 99US-0135280.

(UNMT) UNIV MICHIGAN TECHNOLOGICAL.

Chiang VL, Wu L, Chandrashekhar JP, Carraway DT;

WPI; 2001-032025/04.
P-PSDB; AAB37893.

Novel polynucleotide encoding cellulose synthase useful for enhancing cellulose biosynthesis, modifying lignin biosynthesis, and for accelerating growth of plants and for upregulating cellulose synthase levels -

Claim 1; Fig 1; 65pp; English.

The present sequence is given in a specification relating to polynucleotides encoding cellulose synthase enzymes from Populus tremuloides and Arabidopsis. The cellulose synthases are designated PtCelA and AraxCelA respectively. The polynucleotides may be introduced into plants where they are useful for altering the growth of a plant. Vectors comprising a cellulose synthase promoter, or a fragment of the promoter containing one or more mechanical stress response elements (MSRE), operably linked to a cellulose synthase gene are useful for causing stress-induced gene expression in a plant cell. A cDNA encoding a protein that binds to a positive MSRE may be delivered into the plant to increase expression of cellulose synthase. A cDNA in an antisense orientation may be delivered to reduce expression of cellulose synthase. The polynucleotides enable the development of transgenic tree varieties having increased cellulose content, decreased lignin content and therefore improved wood fibre characteristics. The cellulose synthase promoters are useful for synthesising greater quantities of high-crystalline cellulose in plant and preferably in trees, which permits an increase in proportion of cellulose in transgenic plants, greater strength of juvenile wood and fiber and acceleration of overall growth rate.

Sequence 3232 BP; 895 A; 668 C; 772 G; 897 T; 0 other;

ignment Scores:

ed. No.:	0	Length:	3232
ore:	978.00	Matches:	978
ercent Similarity:	100.00%	Conservative:	0
st Local Similarity:	100.00%	Mismatches:	0
uery Match:	100.00%	Indels:	0
:	22	Gaps:	0

S-09-980-043-2 (1-978) x AAC65448 (1-3232)

```
Qy          1 MetMetGluSerGlyAlaProIleCysHisThrCysGlyGluGlnValGlyHisAspAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         69 ATGATGGAATCTGGGGCTCCTATATGCCATACCTGTGGTGAACAGGTGGGGCATGATGCA 128

Qy         21 AsnGlyGluLeuPheValAlaCysHisGluCysSerTyrProMetCysLysSerCysPhe 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        129 AATGGGGAGCTATTTGTGGCTTGCCATGAGTGTAGCTATCCCATGTGCAAGTCTTGTTTC 188

Qy         41 GluPheGluIleAsnGluGlyArgLysValCysLeuArgCysGlySerProTyrAspGlu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        189 GAGTTTGAAATCAATGAGGGCCGGAAAGTTTGCTTGCGGTGTGGCTCGCCATATGATGAG 248

Qy         61 AsnLeuLeuAspAspValGluLysLysGlySerGlyAsnGlnSerThrMetAlaSerHis 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        249 AACTTGCTGGATGATGTAGAAAAGAAGGGGTCTGGCAATCAATCCACAATGGCATCTCAC 308

Qy         81 LeuAsnAspSerGlnAspValGlyIleHisAlaArgHisIleSerSerValSerThrVal 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        309 CTCAACGATTCTCAGGATGTCGGAATCCATGCTAGACATATCAGTAGTGTGTCCACTGTG 368

Qy        101 AspSerGluMetAsnAspGluTyrGlyAsnProIleTrpLysAsnArgValLysSerCys 120
              ||||:|||:|||  |||:| |||  |||  :|||:|||||||||||, ||||'.'|
Db        369 GATAGTGAAATGAA TATGAATATGGAAATCCAATTTGGAAGAATCGGGTGAAGAGCTGT 428

Qy        121 LysAspLysGluAsnLysLysLysLysArgSerProLysAlaGluThrGluProAlaGln 140
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        429 AAGGATAAAGAGAA AAGAAGAAAAAGAGAAGTCCTAAGGCTGAAACTGAACCAGCTCAA 488

Qy        141 ValProThrGluGl GlnMetGluGluLysProSerAlaGluAlaSerGluProLeuSer 160
              |||||| ||| |:| ||||| |||:| |||.:|||||'||||||:|||||.'||||',
Db        489 GTTCCTACAGAAGA TAGATGGAAGAGAAACCGTCTTGCAGAGGCTTCGGAGCCGCTTTCA 548

Qy        161 IleValTyrProIl ProArgAsnLysLeuThrProTyrArgIleValIleIleMetArg 180
              |||||||:|||:|| |||||||||||||||||||||||||||||||||||||||||||||
Db        549 ATTGTTTATCCAAT TACGTAAGAAAGTCACACCATACAGAGCAGTGATCATTATGCGA 608

Qy        181 LeuValIleLeu  ProLeuGlnGlyPheAsnIleThrAsnProValAspSerAlaPhe 200
              ||  |||: |||  |||||||||||||||||||||||||||||||||||||||||||||
Db        609 CTGGTCATT TTG  TTGCAGGGATTCAACAAATCCTGTTGATATTGCTTT 668

Qy        201 GlyLeuTrpLeuSe SerValIleCysAlaIleTrpPheAlaPheSerTrpValLeuAsp 220
              |||:|||'|||  |||,|||  |||  ||||||||||||||||||||||||||||||||||
Db        669 GGCCTGTGGCTTAG TCTGTCATATGTGAGATCTGGTTTGCATTTTCTTGGGTGTTGGAT 728

Qy        221 GlnPheProLysThr AsnProValAsnArgGluThrTyrIleGluArgLeuSerAlaArg 240
              ||||:|||:|||:|| |||||||||||||||||||||||||||||||||||||||||||||
Db        729 CAGTTCCCCAAGT AAATCCTGTCAATAGAGAAACGTATATCGAAAGGCTGTCGGCAAGG 788

Qy        241 TyrGluArgGluGl GluProSerGlnLeuAlaGlyValAspPheValSerThrVal 260
              |:||||:|||||||  |||||:|||:|.|||:|||||||||||:|||:|||'.|||',
Db        789 TATGAAAGAGAGGT GAGCCTTCTCAGCTTGCCGGTGTGGATTTTTTTGTGAGTACTGTT 848

Qy        261 AspProLeuLysGl ProProLeuIleThrAlaAsnThrValLeuSerIleLeuAlaVal 280
              |||||||||||||  |:|:|||||||||:|||:|||||:|||:|||||.|:|||  ||||.'
Db        849 GATCCGCTGAAGGA CCGCCATTGATCACTGCTAATACAGTCCTTTCCATCTTGCTGTG 908

Qy        281 AspTyrProValAs LysValSerCysTyrValSerAspAspGlyAlaAlaMetLeuSer 300
              ||||||||||||||  |||  |||||||||||||||||||||||||||||||||||||||
Db        909 GACTATCCCGTCGA AAAGTCTCCTGCTACGTGTCTGATGATGGTGCAGCTATGCTTTCA 968

Qy        301 PheGluSerLeuVa GluThrAlaGluPheAlaArgLysTrpValProPheCysLysLys 320
              ||| |||:|||:  |||:|||:  |||  |||||:|||||  |||| | |||
Db        969 TTTGAATCTCTTGT AGAAACAGCTGAGTTTGCAAGGAAGTGGGTTCCGTTCTGCAAAAAA 1028

Qy        321 PheSerIleGluPr ArgAlaProGluPheTyrPheSerGlnLysIleAspTyrLeuLys 340
              ||||:||||:|||  |||||:|||||  |||:|||||||||||||||||:|||||:|||
Db       1029 TTCTCAATTGAACC AAGAGCACCCGAGTTTTACTTCTCACAGAAAATTGATTACTTGAAA 1088

Qy        341 AspLysValGlnPr SerPheValLysGluArgArgAlaMetLysArgAspTyrGluGlu 360
              |||:|||:|||  |||||||||  |||  ||||:|||||||||||||||||||:|||||
Db       1089 GACAAGGTTCAACC TTCTTTGTGAAAGAACGTAGAGCAATGAAAAGGGATTATGAAGAG 1148

Qy        361 TyrLysValArgVa AsnAlaLeuValAlaLysAlaGlnLysThrProGluGluGlyTrp 380
              |||||'|||'|||  |||:|||'|||  |||'|||'|||||||||||||||:|||||||
Db       1149 TACAAAGTCCGAGT AATGCCCTGGTAGCAAAGGCTCAGAAAACACCTGAAGAAGGATGG 1208

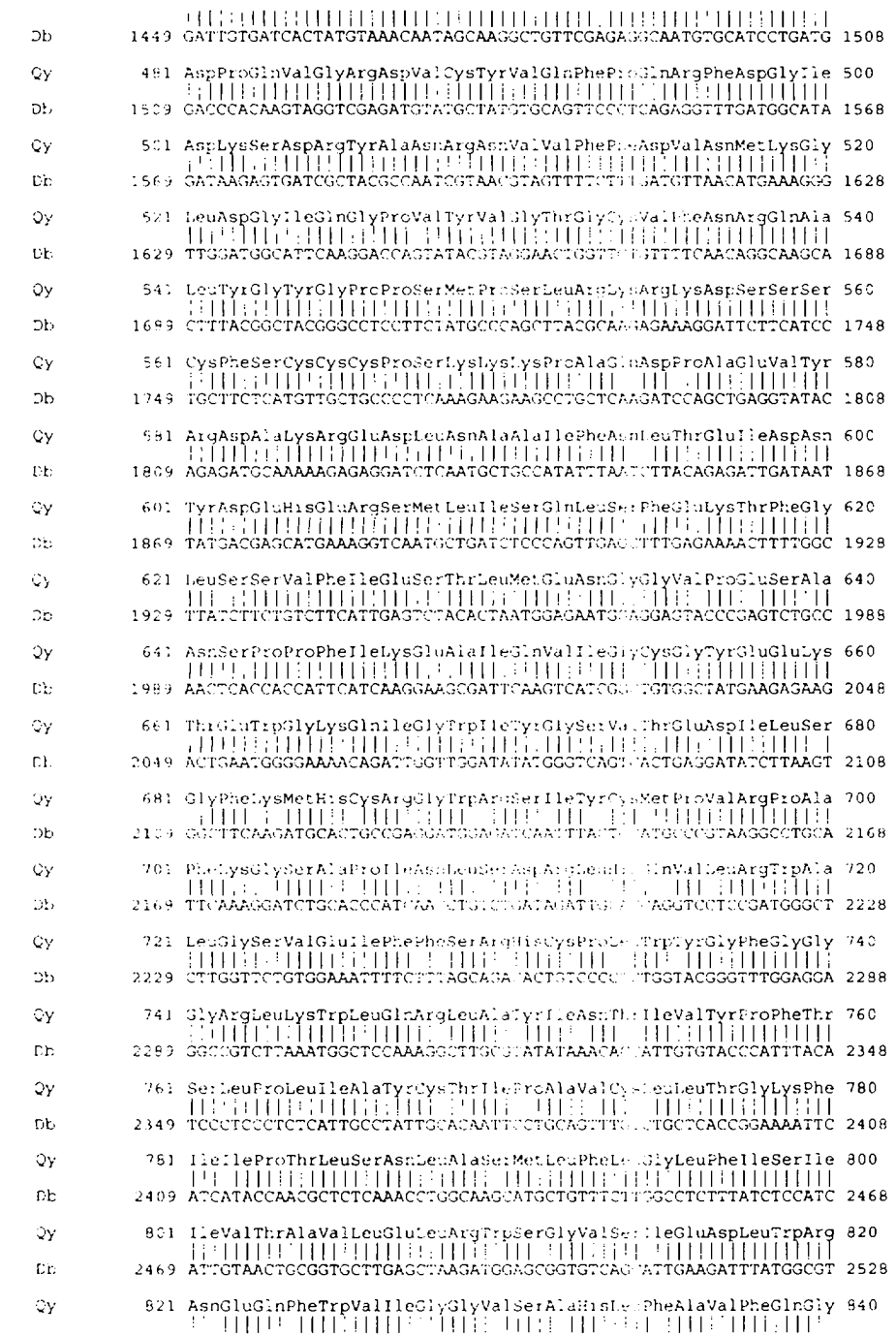
Qy        381 ThrMetGlnAspGl ThrProTrpProGlyAsnAsnThrArgAspHisProGlyHisAsp 400
              ||||||:|||||'|  |||:|'|||:|||||||||||||||||||||||||||||'||
Db       1209 ACTATGCAAGATG TAATACCTTGGCCTGGGAACAACACACGTGATCACCCTGGCCATGAT 1268

Qy        401 SerGlyLeuProTr GluIleLeuGlyAlaArgAspIleGluGlyAsnGluLeuProArg 420
              |||||:|||:|||  |||:|||||||||||||||||||||||||||||||||||||||||
Db       1269 TCAGGTCTTCCTTG GAAATACTGGGAGCTCGTGACATTGAAGGAAATGAACTACCTCGT 1328

Qy        421 LeuValTyrValSe ArgGluLysArgProGlyTyrGlnHisHisLysLysAlaGlyAla 440
              |||||||.||||  |||'||||  |||:|||:|||||||||||:|||||||:|||||||
Db       1329 CTAGTATATGTCTC AGGGAGAAGAGACCTGGCTACCAGCACCACAAAAAGGCTGGTGCA 1388

Qy        441 GluAsnAlaLeuValArgValSerAlaValLeuThrAsnAlaProTyrIleLeuAsnVal 460
              |||||||||||||||||||||||||||||||||:|||||||||||||||||||||:|||  ||
Db       1389 GAAAATGCTCTGGTGAGAGTGTCTGCAGTACTCACAAATGCTCCCTACATCCTCAATGTT 1448

Qy        461 AspCysAspHisTyrValAsnAsnSerLysAlaValArgGluAlaMetCysIleLeuMet 480
```



```
Db       1449 GATTGTGATCACTATGTAAACAATAGCAAGGCTGTTCGAGAGGCAATGTGCATCCTGATG 1508

Qy        481 AspProGlnValGlyArgAspValCysTyrValGlnPheProGlnArgPheAspGlyIle 500

Db       1509 GACCCACAAGTAGGTCGAGATGTATGCTATTTGCAGTTCCCTCAGAGGTTTGATGGCATA 1568

Qy        501 AspLysSerAspArgTyrAlaAsnArgAsnValValPheProAspValAsnMetLysGly 520

Db       1569 GATAAGAGTGATCGCTACGCCAATCGTAACGTAGTTTTCTTTGATGTTAACATGAAAGGG 1628

Qy        521 LeuAspGlyIleGlnGlyProValTyrValGlyThrGlyCysValPheAsnArgGlnAla 540

Db       1629 TTGGATGGCATTCAAGGACCAGTATACGTAGGAACTGGTTGTGTTTCAACAGGCAAGCA 1688

Qy        541 LeuTyrGlyTyrGlyProProSerMetProSerLeuArgLysArgLysAspSerSerSer 560

Db       1689 CTTTACGGCTACGGGCCTCCTTCTATGCCCAGCTTACGCAAAAGAAAGGATTCTTCATCC 1748

Qy        561 CysPheSerCysCysCysProSerLysLysLysProAlaGlnAspProAlaGluValTyr 580

Db       1749 TGCTTCTCATGTTGCTGCCCCTCAAAGAAGAAGCCTGCTCAAGATCCAGCTGAGGTATAC 1808

Qy        581 ArgAspAlaLysArgGluAspLeuAsnAlaAlaIlePheAsnLeuThrGluIleAspAsn 600

Db       1809 AGAGATGCAAAAAGAGAGGATCTCAATGCTGCCATATTTAATCTTACAGAGATTGATAAT 1868

Qy        601 TyrAspGluHisGluArgSerMetLeuIleSerGlnLeuSerPheGluLysThrPheGly 620

Db       1869 TATGACGAGCATGAAAGGTCAATGCTGATCTCCCAGTTGAGCTTTGAGAAAACTTTTGGC 1928

Qy        621 LeuSerSerValPheIleGluSerThrLeuMetGluAsnGlyGlyValProGluSerAla 640

Db       1929 TTATCTTCTGTCTTCATTGAGTCTACACTAATGGAGAATGGAGGAGTACCCGAGTCTGCC 1988

Qy        641 AsnSerProProPheIleLysGluAlaIleGlnValIleGlyCysGlyTyrGluGluLys 660

Db       1989 AACTCACCACCATTCATCAAGGAAGCGATTCAAGTCATCGGCTGTGGCTATGAAGAGAAG 2048

Qy        661 ThrGluTrpGlyLysGlnIleGlyTrpIleTyrGlySerValThrGluAspIleLeuSer 680

Db       2049 ACTGAATGGGGAAAACAGATTGGTTGGATATATGGGTCAGTTACTGAGGATATCTTAAGT 2108

Qy        681 GlyPheLysMetHisCysArgGlyTrpArgSerIleTyrCysMetProValArgProAla 700

Db       2109 GGCTTCAAGATGCACTGCCGAGGATGGAGATCAAGTTATTTTATGCCGTAAGGCCTGCA 2168

Qy        701 PheLysGlySerAlaProIleAsnLeuSerAspArgLeuHisGlnValLeuArgTrpAla 720

Db       2169 TTCAAAGGATCTGCACCCATTAACCTGTCTGATAGATTGCATCAGGTCCTCCGATGGGCT 2228

Qy        721 LeuGlySerValGluIlePhePheSerArgHisCysProLeuTrpTyrGlyPheGlyGly 740

Db       2229 CTTGGTTCTGTGGAAATTTTCTTTAGCAGAACTGTCCCCATTGGTACGGGTTTGGAGGA 2288

Qy        741 GlyArgLeuLysTrpLeuGlnArgLeuAlaTyrIleAsnThrIleValTyrProPheThr 760

Db       2289 GGCGGTCTTAAATGGCTCCAAAGGCTTGCGTATATAAACACTATTGTGTACCCATTTACA 2348

Qy        761 SerLeuProLeuIleAlaTyrCysThrIleProAlaValCysLeuLeuThrGlyLysPhe 780

Db       2349 TCCCTCCCTCTCATTGCCTATTGCACAATTCCTGCAGTTTCTGCTCACCGGAAAATTC 2408

Qy        781 IleIleProThrLeuSerAsnLeuAlaSerMetLeuPheLeuGlyLeuPheIleSerIle 800

Db       2409 ATCATACCAACGCTCTCAAACCTGGCAAGCATGCTGTTTCTTGGCCTCTTTATCTCCATC 2468

Qy        801 IleValThrAlaValLeuGluLeuArgTrpSerGlyValSerIleGluAspLeuTrpArg 820

Db       2469 ATTGTAACTGCGGTGCTTGAGCTAAGATGGAGCGGTGTCAGTATTGAAGATTTATGGCGT 2528

Qy        821 AsnGluGlnPheTrpValIleGlyGlyValSerAlaHisLeuPheAlaValPheGlnGly 840

Db       2529 AATGAACAATTCTGGGTGATCGGAGGTGTTTCAGCCCATCTCTTTGCGGTCTTCCAGGGA 2588

Qy        841 PheLeuLysMetLeuAlaGlyIleAspThrAsnPheThrValThrAlaLysAlaAlaGlu 860

Db       2589 TTCTTAAAAATGTTGGCTGGCATCGATACGAACTTCACTGTCACAGCAAAAGCAGCCGAA 2648

Qy        861 AspAlaGluPheGlyGluLeuTyrMetValLysTrpThrThrLeuLeuIleProProThr 880

Db       2649 GATGCAGAATTTGGGGAGCTATATATGGTCAAGTGGACAACACTTTTGATTCCTCCAACC 2708

Qy        881 ThrLeuLeuIleIleAsnMetSerGlyCysAlaGlyPheSerAspAlaLeuAsnLysGly 900

Db       2709 ACACTTCTCATTATCAATATGTCGGGTTGTGCTGGATTCTCTGATGCACTCAACAAAGGA 2768

Qy        901 TyrGluAlaTrpGlyProLeuPheGlyLysValPhePheAlaPheTrpValIleLeuHis 920

Db       2769 TATGAAGCATGGGGGCCTCTCTTTGGCAAGGTGTTCTTTGCTTTCTGGGTGATTCTTCAT 2828

Qy        921 LeuTyrProPheLeuLysGlyLeuMetGlyArgGlnAsnLeuThrProThrIleValVal 940

Db       2829 CTCTATCCATTCCTTAAAGGTCTAATGGGTCGCCAAAACCTAACACCAACCATTGTTGTT 2888

Qy        941 LeuTrpSerValLeuLeuAlaSerValPheSerLeuValTrpValLysIleAsnProPhe 960

Db       2889 CTCTGGTCAGTGCTGTTGGCCTCTGTCTTCTCTCTCGTTTGGGTCAAGATCAATCCATTC 2948

Qy        961 ValAsnLysValAspAsnThrLeuValAlaGluThrCysIleSerIleAspCys 978

Db       2949 GTTAACAAAGTTGATAACACCTTGGTTGCGGAGACCTGCATTTCCATTGATTGC 3002

RESULT 2
AAV08372
ID   AAV08372 standard; cDNA to mRNA; 3207 BP.
XX
AC   AAV08372;
XX
DT   08-FEB-1999  (first entry)
XX
DE   Cellulose synthase PcsA1 coding sequence.
XX
KW   Cellulose synthase; cotton; ds.
XX
OS   Gossypium hirsutum.
XX
FH   Key             Location/Qualifiers
FT   CDS             77..3001
FT                   /*tag= a
XX
PN   EP876575 A2
XX
PD   04-NOV-1998.
XX
PF   31-MAR-1998;   98EP-0302489.
XX
PR   01-APR-1997;   97JP-0083133.
XX
PA   (NISN ) NISSHINBO IND INC.
XX
PI   Aotsuka S,  Hasegawa O,  Hayashi T,  Ihara Y;
XX
DR   WPI; 1998-559440/48.
DR   P-PSDB; AAW73308.
XX
PT   New DNA encoding cotton cellulose synthase proteins - useful for
PT   controlling cellulose synthesis in prokaryotic or eukaryotic cells
XX
PS   Example 5; Page 14-18; 51pp; English.
XX
CC   This sequence encodes a cellulose synthase of the invention. The DNA can
CC   be used for controlling cellulose synthesis in a cell by introducing the
CC   DNA into the cell, expressing RNA with a nucleotide sequence homologous
CC   to the DNA or a nucleotide sequence complementary to the DNA. The cells
CC   are prokaryotic cells i.e. acetobacterium and/or eukaryotic cells such as
```

yeasts i.e. Saccharomyces spp., plant cells (cotton plants) and cultured cells of mammals etc.

Sequence 3207 BP; 883 A; 662 C; 723 G; 939 T; 0 other;

gnment Scores:

ed. No.:	1.05e-52	Length:	3207
ore:	61.00	Matches:	61
rcent Similarity:	100.00%	Conservative:	0
st Local Similarity:	100.00%	Mismatches:	0
ery Match:	6.24%	Indels:	0
	19	Gaps:	0

-09-980-043-2 (1-978) x AAV08372 (1-3207)

```
 799 SerIleIleValThrAlaValLeuGluLeuArgTrpSerGlyValSerIleGluAspLeu 818
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
2456 TCCATTATCGTGACTGCTGTTCTCGAGCTCCGATGGAGTGGTGTCAGCATTGAGGACTTA 2515

 819 TrpArgAsnGluGlnPheTrpValIleGlyGlyValSerAlaHisLeuPheAlaValPhe 838
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
2516 TGGCGTAACGAGCAGTTTTGGGTCATCGGTGGCGTTTCAGCCCATCTCTTTGCCGTCTTC 2575

 839 GlnGlyPheLeuLysMetLeuAlaGlyIleAspThrAsnPheThrValThrAlaLysAla 858
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
2576 CAAGGTTTCCTTAAGATGCTTGCGGGCATTGACACCAACTTTACTGTCACTGCCAAAGCA 2635

 859 Ala 859
     |||
2636 GCT 2638
```

SULT 3
V34432
AAV34432 standard; cDNA; 3328 BP.

AAV34432;

28-SEP 1998 (first entry)

Cotton cellulose synthase CelA1 cDNA.

Cellulose synthase; CelA1 gene; cotton; forest tree; plant fibre; wood; transgenic plant; ds.

Gossypium hirsutum Acala SJ-2.

Key	Location/Qualifiers
CDS	179..3103
	/*tag= a

WO9818949-A2.

07-MAY-1998.

29-OCT-1997; 97WO-US19529.

29-OCT-1996; 96US-0029987.

(CALJ) CALGENE INC.

Delmer DP, Pear JR, Stalker DM;

WPI; 1998-272238/24.
P-PSDB; AAW60039.

New isolated plant cellulose synthesis enzyme DNA - is used to develop products for producing modified plant fibre products, particularly for modifying cotton fibre and wood quality

Claim 5; Fig 6A-E; 49pp; English.

This full-length cDNA sequence corresponds to the cotton CelA1 gene transcript and encodes a 973-amino acid cellulose synthase (see

CC AAW60039). A cDNA clone that exhibited small blocks of amino acid
CC homology to proteins encoded by bacterial CelA genes was isolated
CC from an unamplified cDNA library prepared from mRNA of fibres of
CC cotton Acala SJ-2 harvested at 21 dpa (time of maximal secondary
CC wall cellulose synthesis). These 2 clones were partially homologous
CC at the nucleotide and amino acid levels and were designated CelA1
CC and CelA2 (see AAV34431 and AAV34432, respectively). The 2 cotton CelA genes are
CC both expressed in developing fibre at approximately 500 times their
CC level of expression in other cotton tissues, and constitute
CC approximately 1.2% of the 24 dpa fibre mRNA. The CelA1 gene
CC promoter region (see AAV34433) is also claimed. The isolated cDNA
CC clones are used in claimed methods for modifying fibre phenotype
CC in a cotton plant and for modifying wood quality phenotype in a
CC forest tree species. Fibre strength can be reduced by over-expressing the
CC cellulose synthase protein and diverting production into cellulose.
XX
SQ Sequence 3328 BP; 883 A; 662 C; 723 G; 939 T; 0 other;

Alignment Scores:

Pred. No.:	1.0[illegible]	Matches:	3328
Score:	61.00	Matches:	61
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	6.24%	Indels:	0
DB:	19	Gaps:	0

US 09-980 043-2 (1-978) x AAV34432 (1-3328)

```
Qy        799 SerIleIleValThrAlaValLeuGluLeuArgTrpSerGlyValSerIleGluAspLeu 818
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2558 TCCATTATCGTGACTGCTGTTCTCGAGCTCCGATGGAGTGGTGTCAGCATTGAGGACTTA 2617

Qy        819 TrpArgAsnGluGlnPheTrpValIleGlyGlyValSerAlaHisLeuPheAlaValPhe 838
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2618 TGGCGTAACGAGCAGTTTTGGGTCATCGGTGGCGTTTCAGCCCATCTCTTTGCCGTCTTC 2677

Qy        839 GlnGlyPheLeuLysMetLeuAlaGlyIleAspThrAsnPheThrValThrAlaLysAla 858
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2678 CAAGGTTTCCTTAAGATGCTTGCGGGCATTGACACCAACTTTACTGTCACTGCCAAAGCA 2737

Qy        859 Ala 859
              |||
Db       2738 GCT 2740
```

RESULT 4
AAC65449
ID AAC65449 standard; DNA; 1010 BP.
XX
AC AAC65449;
XX
DT 07-MAR 2001 (first entry)
XX
DE Populus tremuloides CelA 5' flanking region.
XX
KW Aspen; CelA; cellulose synthase; arabidopsis; wood;
KW mechanical stress responsive element; MSRE; cellulose biosynthesis; ds.
XX
OS Populus tremuloides.
XX
PN WO200071670-A2.
XX
PD 30-NOV-2000.
XX
PF 18-MAY 2000; 2000WO-US13637.
XX
PR 21-MAY-1999; 99US-0135280.
XX
PA (UNMT) UNIV MICHIGAN TECHNOLOGICAL.
XX
PI Chiang VL, Wu L, Chandrashekhar JP, Carraway DT;
XX
DR WPI; 2001-032025/04.
DR P-PSDB; AAB37894.

```
XX
PT   Novel polynucleotide encoding cellulose synthase useful for enhancing
PT   cellulose biosynthesis, modifying lignin biosynthesis, and for
PT   accelerating growth of plants and for upregulating cellulose synthase
PT   levels -
XX
PS   Claim 2; Fig 4; 69pp; English.
XX
CC   The present sequence is given in a specification relating to
CC   polynucleotides encoding cellulose synthase enzymes from Populus
CC   tremuloides and Arabidopsis. The cellulose synthases are designated
CC   PtCelA and AraxCelA respectively. The polynucleotides may be introduced
CC   into plants where they are useful for altering the growth of a plant.
CC   Vectors comprising a cellulose synthase promoter, or a fragment of the
CC   promoter containing one or more mechanical stress response elements
CC   (MSRE), operably linked to a cellulose synthase gene are useful for
CC   causing stress induced gene expression in a plant cell. A cDNA encoding a
CC   protein that binds to a positive MSRE may be delivered into the plant to
CC   increase expression of cellulose synthase. A cDNA in an antisense
CC   orientation may be delivered to reduce expression of cellulose synthase.
CC   The polynucleotides enable the development of transgenic tree varieties
CC   having increased cellulose content, decreased lignin content and
CC   therefore improved wood fibre characteristics. The cellulose synthase
CC   promoters are useful for synthesising greater quantities of
CC   high-crystalline cellulose in plant and preferably in trees, which
CC   permits an increase in proportion of cellulose in transgenic plants,
CC   greater strength of juvenile wood and fiber and acceleration of overall
CC   growth rate
XX
SQ   Sequence 1010 BP; 289 A; 212 C; 134 G; 335 T; 0 other;

Alignment Scores:
Pred. No.:              1.69e-35        Length:         1010
Score:                  44.00           Matches:        44
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            4.50%           Indels:         0
DB:                     22              Gaps:           0

US-09-980-043-2 (1-978) x AAC65449 (1..1010)

Qy          1 MetMetGluSerGlyAlaProIleCysHisCysCysGlyGlnValGlyHisAspAla 20
              ||||||||||||||||||||||||||||||||||||||||
Db        841 ATGATGGAATCTGGGGCTCCTATATGTCATGCTGTGGTCAGGTCGGGCATGATGCA 900

Qy         21 AsnGlyGluLeuPheValAlaCysHisGluCysSerTyrProMetCysLysSerCysPhe 40
              ||| |||||||||||| ||| ||| ||| ||| ||| |||||||||
Db        901 AATGGGGAGCTATTTGTGGCTTGTCATGAACTTAATATCATGTCAAGTCTTGTTTC 960

Qy         41 GluPheSerIle 44
              ||| |||||
Db        961 GAGTTTGAAATC 972

RESULT 5
AAC50201
ID   AAC50201 standard; DNA; 950 BP
XX
AC   AAC50201;
XX
DT   18-OCT-2000 (first entry)
XX
DE   Arabidopsis thaliana DNA fragment Seq ID No. 6895.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway;
KW   metabolic pathway; promoter; termination sequence; ds.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
PR   18-JUN-1999;   99US-0139456.
PR   18-JUN-1999;   99US-0139457.
PR   18-JUN-1999;   99US-0139458.
PR   18-JUN-1999;   99US-0139459.
PR   18-JUN-1999;   99US-0139460.
PR   18-JUN-1999;   99US-0139461.
PR   18-JUN-1999;   99US-0139462.
PR   18-JUN-1999;   99US-0139463.
PR   18-JUN-1999;   99US-0139750.
PR   18-JUN-1999;   99US-0139763.
PR   21-JUN-1999;   99US-0139817.
PR   22-JUN-1999;   99US-0139899.
PR   23-JUN-1999;   99US-0140353.
PR   23-JUN-1999;   99US-0140354.
PR   24-JUN-1999;   99US-0140695.
PR   28-JUN-1999;   99US-0140823.
PR   29-JUN-1999;   99US-0140991.
PR   30-JUN-1999;   99US-0141287.
PR   01-JUL-1999;   99US-0141842.
PR   01-JUL-1999;   99US-0142154.
PR   02-JUL-1999;   99US-0142055.
PR   06-JUL-1999;   99US-0142390.
PR   08-JUL-1999;   99US-0142803.
```

09-JUL-1999; 99US-0142920.
12-JUL-1999; 99US-0142977.
13-JUL-1999; 99US-0143542.
14-JUL-1999; 99US-0143624.
15-JUL-1999; 99US-0144005.
16-JUL-1999; 99US-0144085.
16-JUL-1999; 99US-0144086.
19-JUL-1999; 99US-0144325.
19-JUL-1999; 99US-0144331.
19-JUL-1999; 99US-0144332.
19-JUL-1999; 99US-0144333.
19-JUL-1999; 99US-0144334.
19-JUL-1999; 99US-0144335.
20-JUL-1999; 99US-0144352.
20-JUL-1999; 99US-0144632.
20-JUL-1999; 99US-0144884.
21-JUL-1999; 99US-0144814.
21-JUL-1999; 99US-0145086.
21-JUL-1999; 99US-0145088.
22-JUL-1999; 99US-0145085.
22-JUL-1999; 99US-0145087.
22-JUL-1999; 99US-0145089.
22-JUL-1999; 99US-0145192.
23-JUL-1999; 99US-0145145.
23-JUL-1999; 99US-0145218.
23-JUL-1999; 99US-0145224.
26-JUL-1999; 99US-0145276.
27-JUL-1999; 99US-0145913.
27-JUL-1999; 99US-0145918.
27-JUL-1999; 99US-0145919.
28-JUL-1999; 99US-0145951.
02-AUG-1999; 99US-0146386.
02-AUG-1999; 99US-0146388.
02-AUG-1999; 99US-0146389.
03-AUG-1999; 99US-0147038.
04-AUG-1999; 99US-0147204.
04-AUG-1999; 99US-0147302.
05-AUG-1999; 99US-0147192.
05-AUG-1999; 99US-0147260.
06-AUG-1999; 99US-0147303.
06-AUG-1999; 99US-0147416.
09-AUG-1999; 99US-0147493.
09-AUG-1999; 99US-0147935.
10-AUG-1999; 99US-0148171.
11-AUG-1999; 99US-0148319.
12-AUG-1999; 99US-0148341.
13-AUG-1999; 99US-0148565.
13-AUG-1999; 99US-0148684.
16-AUG-1999; 99US-0149368.
17-AUG-1999; 99US-0149175.
18-AUG-1999; 99US-0149426.
20-AUG-1999; 99US-0149722.
20-AUG-1999; 99US-0149723.
20-AUG-1999; 99US-0149929.
23-AUG-1999; 99US-0149902.
23-AUG-1999; 99US-0149930.
25-AUG-1999; 99US-0150566.
26-AUG-1999; 99US-0150884.
27-AUG-1999; 99US-0151065.
27-AUG-1999; 99US-0151066.
27-AUG-1999; 99US-0151080.
30-AUG-1999; 99US-0151303.
31-AUG-1999; 99US-0151438.
01-SEP-1999; 99US-0151930.
07-SEP-1999; 99US-0152363.
10-SEP-1999; 99US-0153070.
13-SEP-1999; 99US-0153758.
15-SEP-1999; 99US-0154018.
16-SEP-1999; 99US-0154039.
20-SEP-1999; 99US-0154779.
22-SEP-1999; 99US-0155139.
23-SEP-1999; 99US-0155486.
24-SEP-1999; 99US-0155659.

PR 28-SEP-1999; 99US-0156459.
PR 29-SEP-1999; 99US-0156796.
PR 04-OCT-1999; 99US-0158117.
PR 05-OCT-1999; 99US-0158753.
PR 06-OCT-1999; 99US-0159065.
PR 07-OCT-1999; 99US-0159429.
PR 08-OCT-1999; 99US-0159832.
PR 12-OCT-1999; 99US-0159966.
PR 13-OCT-1999; 99US-0160293.
PR 13-OCT-1999; 99US-0160794.
PR 13-OCT-1999; 99US-0160895.
PR 14-OCT-1999; 99US-0161329.
PR 14-OCT-1999; 99US-0161430.
PR 14-OCT-1999; 99US-0161481.
PR 14-OCT-1999; 99US-0161837.
PR 14-OCT-1999; 99US-0161938.
PR 18-OCT-1999; 99US-0162564.
PR 21-OCT-1999; 99US-0163111.
PR 21-OCT-1999; 99US-0163[illegible].
PR 21-OCT-1999; 99US-0163[illegible].
PR 21-OCT-1999; 99US-0163[illegible].
PR 21-OCT-1999; 99US-0163814.
PR 21-OCT-1999; 99US-0163915.
PR 22-OCT-1999; 99US-0164080.
PR 22-OCT-1999; 99US-0164081.
PR 22-OCT-1999; 99US-0164089.
PR 25-OCT-1999; 99US-0164404.
PR 25-OCT-1999; 99US-0164405.
PR 25-OCT-1999; 99US-0164406.
PR 26-OCT-1999; 99US-0164659.
PR 26-OCT-1999; 99US-0164860.
PR 26-OCT-1999; 99US-0164861.
PR 28-OCT-1999; 99US-0165920.
PR 28-OCT-1999; 99US-0165992.
PR 28-OCT-1999; 99US-0166093.
PR 29-OCT-1999; 99US-0166242.

Alignment Scores:
Pred. No.: 1.75e-44 Length: 950
Score: 483.0 Matches: 43
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 4.48% Indels: 0
DB: 21 Gaps: 0

US-09-980-043-2 (1-978) x AAC50231 (1-960)

Qy 805 ValLeuGluLeuAspTrpSerGlyValSerIleGluAspLeuTrpArgAsnGluGlnPhe 824
Db 366 GTCCTCGAGCTTGAATGGAGCGGAGTCAGTATCGAAGACTTATGGAGAAACGAACAGTTT 425

Qy 825 TrpValIleGlyGlyValSerAlaAsnLeuProAlaValPheGlnGlyPheLeuLysMet 844
Db 426 TGGGTTATTGGAGGAGTCTCAGCTCAACTTCCTGCTGTTTTCCAAGGATTCCTCAAAATG 485

Qy 845 LeuAlaGly 853
Db 486 CTCGCTGGT 494

RESULT 6
AAC33457
ID AAC33457 standard; DNA; 951 BP.
XX
AC AAC33457;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 3109.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX

OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138034.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.

15-SEP-1999; 99US-0154018.
16-SEP-1999; 99US-0154039.
20-SEP-1999; 99US-0154779.
22-SEP-1999; 99US-0155139.
23-SEP-1999; 99US-0155486.
24-SEP-1999; 99US-0155659.
28-SEP-1999; 99US-0156458.
29-SEP-1999; 99US-0156596.
04-OCT-1999; 99US-0157117.
05-OCT-1999; 99US-0157753.
06-OCT-1999; 99US-0157865.
07-OCT-1999; 99US-0158029.
08-OCT-1999; 99US-0158232.
12-OCT-1999; 99US-0158369.
13-OCT-1999; 99US-0159293.
13-OCT-1999; 99US-0159294.
13-OCT-1999; 99US-0159295.
14-OCT-1999; 99US-0159329.
14-OCT-1999; 99US-0159330.
14-OCT-1999; 99US-0159331.
14-OCT-1999; 99US-0159637.
14-OCT-1999; 99US-0159638.
18-OCT-1999; 99US-0159584.
21-OCT-1999; 99US-0160741.
21-OCT-1999; 99US-0160767.
21-OCT-1999; 99US-0160768.
21-OCT-1999; 99US-0160770.
21-OCT-1999; 99US-0160814.
21-OCT-1999; 99US-0160815.
22-OCT-1999; 99US-0160980.
22-OCT-1999; 99US-0160981.
22-OCT-1999; 99US-0160989.
25-OCT-1999; 99US-0161404.
25-OCT-1999; 99US-0161405.
25-OCT-1999; 99US-0161406.
26-OCT-1999; 99US-0161359.
26-OCT-1999; 99US-0161360.
26-OCT-1999; 99US-0161361.
28-OCT-1999; 99US-0161920.
28-OCT-1999; 99US-0161992.
28-OCT-1999; 99US-0161993.
29-OCT-1999; 99US-0162142.

ignment Scores:
ed. No.: 1.75e-34 Length: 951
ore: 43.00 Matches: 43
rcent Similarity: 100.00% Conservative: 0
st Local Similarity: 100.00% Mismatches: 0
ery Match: 4.40% Indels: 0
: 21 Gaps: 0

-09-980-043-2 (1-978) x AAC33457 (1-951)

805 ValLeuGluLeuArgTrpSerGlyValSerIleGluAspLeuTrpArgAsnGluGlnPhe 824
368 GTCCTCGAGCTTCGATGGAGCGGAGTCAGTATCGAAGACTTATGGAGAAACGAACAGTTT 427

825 TrpValIleGlyGlyValSerAlaHisLeuPheAlaValPheGlnGlyPheLeuLysMet 844
428 TGGGTTATTGGAGGTGTCTCAGCTCATCTCTTTGCCGTTTTCCAAGGATTCCTCAAAATG 487

845 LeuAlaGly 847
488 CTAGCTGGT 496

SULT 7
A67145
AAA67145 standard; DNA; 3747 BP.

AAA67145;

31-OCT-2000 (first entry)

DE Eucalyptus grandis cell wall synthase nucleotide sequence SEQ ID NO:141.
XX
KW Eucalyptus grandis; Pinus radiata; Monterey pine; modification;
KW plant cell wall; polysaccharide; polysaccharide biosynthetic pathway;
KW transgenic plant; ds.
XX
OS Eucalyptus grandis
XX
PN WO200022092-A2.
XX
PD 20-APR-2000.
XX
PF 08-OCT-1999; 99WO-NZ [illegible].
XX
PR 13-OCT-1998; 98US-01[illegible].
PR 11-AUG-1999; 99US-01[illegible].
XX
PA (GENE-) GENESIS PLC.
PA (FLET-) FLETCHER CHALL[illegible]
XX
PI Bloksberg LN;
XX
DR WPI; 2000-339328/29.
DR P-PSDB; AAB16338.
XX
PT New genes encoding proteins involved in a plant polysaccharide
PT biosynthetic pathway, useful for modulating or altering the
PT polysaccharide content, composition or structure of the plant
XX
PS Claim 1; Page 106-107; [illegible]pp; English
XX
CC The present invention describes isolated polynucleotides (PN) comprising
CC a sequence selected from one of 835 nucleotide sequences given in
CC AAA67073 to AAA67907, their (reverse) complements, sequences producing
CC an Expectation (E) value of 0.01 or less compared to the 835 sequences,
CC sequences at least 50% identical to them, 200, 100, 40 or 20 mers of the
CC 835 sequences or sequences that are degenerately equivalent or allelic
CC to the 835 sequences. The polynucleotides are used to modify the
CC activity of a polypeptide involved in a polysaccharide biosynthetic
CC pathway in the plant. They are especially used to modulate or alter the
CC polysaccharide content, composition or structure of the plant. AAB16268
CC to AAB16340 are proteins encoded by some of the polynucleotide sequence
CC given in the present invention
XX
SQ Sequence 3747 BP; 976 A; 822 C; 935 G; 1014 T; 0 other;

Alignment Scores:
Pred. No.: 6.44e-34 Length: 3747
Score: 43.00 Matches: 43
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.47% Indels: 0
DB: 21 Gaps: 0

US-09-980-043-2 (1-978) x AAA67145 (1-3747)

Qy 891 AlaGlyPheSerAsp[illegible]AlaLeuAsnLysGlyTyr[illegible]GluAlaTrpGlyProLeuPheGlyLys 910
Db 3184 GCCGGGTTCTCCGAC[illegible]GCCTGAA[illegible]AAAGGTTA[illegible]GAGGCGTGGGGACCCCTTTTGGCAAG 3243

Qy 911 ValPhePheAlaPhe[illegible]TrpVal[illegible]eLeuHisLeuTyr[illegible]ProPheLeuLysGlyLeuMetGly 930
Db 3244 GTCTTCTTTGCATT[illegible]GGGTGATT[illegible]TTCATC[illegible]TATCCATTCCTGAAAGGTCTCATGGGT 3303

Qy 931 ArgGlnAsn 933
Db 3304 AGGCAGAAC 3312

RESULT 8
AAA67084
ID AAA67084 standard; DNA; 719 BP.
XX
AC AAA67084;

```
XX
DT   31-OCT-2000  (first entry)
XX
DE   Eucalyptus grandis cellulose synthase nucleotide sequence SEQ ID NO:12.
XX
KW   Eucalyptus grandis; pinus radiata; Monterey pine; modification;
KW   plant cell wall; polysaccharide; polysaccharide biosynthetic pathway;
KW   transgenic plant; ds.
XX
OS   Eucalyptus grandis.
XX
PN   WO200022092 A2.
XX
PD   20-APR-2000
XX
PF   06-OCT-1999;   99WO NZ00169
XX
PR   13-OCT-1998;   98US-0170862.
PR   11-AUG-1999;   99US-0148426.
XX
PA   (GENE-) GENESIS RES & DEV CORP LTD.
PA   (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI   Bicksberg LN;
XX
DR   WPI; 2000-339328/29.
DR   P-PSDB; AAB16279.
XX
PT   New genes encoding proteins involved in a plant polysaccharide
PT   biosynthetic pathway, useful for modulating or altering the
PT   polysaccharide content, composition or structure of the plant -
XX
PS   Claim 1; Page 42; 301pp; English.
XX
CC   The present invention describes isolated polynucleotides (PN) comprising
CC   a sequence selected from one of 835 nucleotide sequences given in
CC   AAA67073 to AAA67907, their (reverse) complements, sequences producing
CC   an Expectation (E) value of 0.01 or less compared to the 835 sequences,
CC   sequences at least 50% identical to them, 200, 150, 40 or 20-mers of the
CC   835 sequences or sequences that are degenerately equivalent or allelic
CC   to the 835 sequences. The polynucleotides are used to modify the
CC   activity of a polypeptide involved in a polysaccharide biosynthetic
CC   pathway in the plant. They are especially used to modulate or alter the
CC   polysaccharide content, composition or structure of the plant. AAB16268
CC   to AAB16340 are proteins encoded by some of the polynucleotide sequence
CC   given in the present invention.
XX
SQ   Sequence 719 BP; 197 A; 147 C; 142 G; 233 T; 0 other;

Alignment Scores:
Pred. No.:                1.75e-31        Length:                 719
Score:                    40.00           Matches:                40
Percent Similarity:       100.00%         Conservative:           0
Best Local Similarity:    100.00%         Mismatches:             0
Query Match:              4.09%           Indels:                 0
DB:                       21              Gaps:                   0

US-09-980-043-2 (1-978) x AAA67084 (1-719)

Qy        256 PheValSerThrValAspProLeuLysGluProProLeuIleThrAlaAsnThrValLeu 275
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         56 TTTGTGAGTACTGTGGATCCTTTAAAGAACCTCCTCTGATAACAGCGAACACTGTCTTA 117

Qy        276 SerIleLeuAlaValAspTyrProValAspLysValSerCysTyrValSerAspAspGly 295
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        118 TCCATACTTGCCGTGGATTACCCGGTCGACAAAGTGTCGTGCTATGTCTCTGATGATGGG 177

RESULT 9
AAV06569
ID   AAV06569 standard; cDNA; 1741 BP
XX
AC   AAV06569;
XX
DT   06-JUL-1998  (first entry)
XX
DE   Rice cellulose synthase RSW1-like cDNA clone S0542.
XX
KW   Cellulose synthase; RSW1-like gene; beta-1,4-glucan;
KW   transgenic plant; rice; ss.
XX
OS   Oryza sativa.
XX
FH   Key             Location/Qualifiers
FT   CDS             101..1741
FT                   /*tag= a
XX
PN   WO9800549-A1.
XX
PD   08-JAN-1998.
XX
PF   24-JUN-1997;   97WO-AU00402.
XX
PR   27-JUN-1996;   96AU-0000699.
XX
PA   (CSIR ) COMMONWEALTH SCI & IND RES ORG.
PA   (AUSU ) UNIV AUSTRALIAN NAT.
XX
PI   Arioli A,  Betzner AS,  Peng L,  Williamson RE;
XX
DR   WPI; 1998-086974/08.
DR   P-PSDB; AAW46202.
XX
PT   DNA encoding cellulose biosynthetic enzyme - useful for manipulation
PT   of cellulose and beta-1,4-glucan
XX
PS   Claim 13; Page 163-166; 207pp; English.
XX
CC   Rice partial cDNA clone EST S0542 was isolated on the basis of
CC   homology to newly identified Arabidopsis cellulose synthase RSW1
CC   nucleotide sequences (see AAV06562-68). It encodes the first 547
CC   amino acids (see AAW46202) of the rice RSW1 or RSW1-like polypeptide,
CC   showing 70% identity to the Arabidopsis RSW1 polypeptide given in
CC   AAW33817. Claimed nucleic acid molecules (see AAV06562-69) coding for
CC   claimed polypeptides (see AAW33816-20 and AAW46202) involved in
CC   cellulose biosynthesis can be used to manipulate the cellulose
CC   and/or beta-glucan content of transgenic plants. Expression of
CC   nucleic acids in the sense orientation increases the level of
CC   cellulose and reduces the level of non-crystalline beta-1,4-glucan
CC   and starch, providing plants with modified strength and/or shape
CC   and/or fibre properties, or having increased resistance to stresses
CC   or pests. Antisense, ribozyme or co-suppression molecules can be
CC   used to reduce the cellulose content of a transgenic plant, e.g. to
CC   improve digestibility or to alter carbon partitioning such that
CC   increased carbon is available for growth.
XX
SQ   Sequence 1741 BP; 440 A; 399 C; 486 G; 416 T; 0 other;

Alignment Scores:
Pred. No.:                4.11e-31        Length:                 1741
Score:                    40.00           Matches:                40
Percent Similarity:       100.00%         Conservative:           0
Best Local Similarity:    100.00%         Mismatches:             0
Query Match:              4.09%           Indels:                 0
DB:                       19              Gaps:                   0

US-09-980-043-2 (1-978) x AAV06569 (1-1741)

Qy        256 PheValSerThrValAspProLeuLysGluProProLeuIleThrAlaAsnThrValLeu 275
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1160 TTTGTCAGTACGGTGGATCCACTAAAGGAACCTCCTCTGATCACAGCAAACACTGTTTTG 1219

Qy        276 SerIleLeuAlaValAspTyrProValAspLysValSerCysTyrValSerAspAspGly 295
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1220 TCCATTCTGGCTGTGGATTACCCTGTTGACAAAGTGTCATGCTATGTTTCTGACGATGGT 1279

RESULT 10
```

99500
AAZ99500 standard; DNA; 3725 BP.

AAZ99500;

03-JUL-2000 (first entry)

DNA encoding a maize cellulose synthase.

Maize; cellulose synthase; stalk quality; stand; silage; cellulose; transgenic plant; plant breeding marker; ss.

Zea mays.

Key	Location/Qualifiers
CDS	179..3398
	/*tag= a
	/product= "cellulose synthase"
	/note= "no termination codon given"

WO200009706 A2.

24-FEB-2000.

16-AUG-1999; 99WO-US18760.

17-AUG-1998; 98US-0096822.

(PION-) PIONEER HI-BRED INT INC.

Dhugga KS, Helentjaris TG, Bowen BA, Wang X;

WPI; 2000-224343/19.
P-PSDB; AAY84110.

New genes which encode maize cellulose synthase polypeptides in plants useful for modulating the expression of cellulose synthase in plants and to produce transgenic plants expressing the novel protein -

Claim 1; Page 105-110; 119pp; English.

The present sequence encodes a maize cellulose synthase polypeptide. The cellulose synthase can be used for the improvement of stalk quality for improved stand or silage. It also provides an increased concentration of cellulose in the pericarp, hardening the kernel and improving its handling ability. The sequences are used to produce transgenic plants and seeds expressing the cellulose synthase. The polynucleotide is used for modulating, preferably increasing, the level of the synthase in a plant cell. The plants are preferably monocots. The polynucleotide is also used as a probe or primer in the detection quantitation or isolation of gene transcripts. The probes are useful in detecting deficiencies in the level of mRNA in screenings for desired transgenic plant, for detecting mutations in the gene, for monitoring upregulation of expression or changes in enzyme activity in screening assays of compounds, for detection of any number of allelic variants of the gene, or for use as molecular markers in plant breeding programs. The isolated nucleic acids of the present invention can also be used for recombinant expression of their encoded polypeptides or for use as immunogens in the preparation and/or screening of antibodies. The proteins can be employed in assays for enzyme agonists or antagonists of enzyme function or for use of immunogens or antigens to obtain antibodies specifically immunoreactive with a protein.

Sequence 3725 BP; 927 A; 807 C; 950 G; 1041 T; 0 other;

ignment Scores:			
ed. No.:	8.45e-31	Length:	3725
ore:	40.00	Matches:	40
rcent Similarity:	100.00%	Conservative:	0
st Local Similarity:	100.00%	Mismatches:	0
ery Match:	4.09%	Indels:	0
:	21	Gaps:	0

US-09-980-043-2 (1-9781) x AA [illegible]

Qy 256 PheValSerThrV... AspProGlu...GluPr...ProLeuIleThrAlaAsnThrValLeu 275
Db 1232 TTTGTCAGTACAG... GATCCATTGAAAGAACCT...ACTGATCACAGCCAACACTGTTTTG 1291

Qy 276 SerIleLeuAlaVa...AspTyrProValAsp...ValSerCysTyrValSerAspAspGly 295
Db 1292 TCCATTCTTGCTT... GATTACCCT...AAAG...GTCATGCTATGTTTCTGATGATGGC 1351

RESULT 11
AAZ99506
ID AAZ99506 standard; DNA [illegible] BP
XX
AC AAZ99506;
XX
DT 03-JUL-2000 (first [illegible]
XX
DE DNA encoding a maize [illegible] gene
XX
KW Maize; cellulose synthase; stalk quality; stand; silage; cellulose; transgenic plant; plant breeding marker; ss.
XX
OS Zea mays.
XX
FH Key Location/Qualifiers
FT CDS 179..3398
FT /*tag= a
FT /product= "cellulose synthase"
FT /note= "no termination codon given"
XX
PN WO200009706 A2.
XX
PD 24-FEB-2000.
XX
PF 16-AUG-1999; 99WO-US18760.
XX
PR 17-AUG-1998; 98US-0096822.
XX
PA (PION-) PIONEER HI-BRED INT INC.
XX
PI Dhugga KS, Helentjaris TG, Bowen BA, Wang X;
XX
DR WPI; 2000-224343/19.
DR P-PSDB; AAY84112.
XX
PT New genes which encode maize cellulose synthase polypeptides in plants
PT useful for modulating the expression of cellulose synthase in plants
PT and to produce transgenic plants expressing the novel protein
XX
PS Claim 1; Page 121-126; 119pp; English.
XX
CC The present sequence encodes a maize cellulose synthase polypeptide.
CC The cellulose synthase can be used for the improvement of stalk quality
CC for improved stand or silage. It also provides an increased concentration
CC of cellulose in the pericarp, hardening the kernel and improving its
CC handling ability. The sequences are used to produce transgenic plants
CC and seeds expressing the cellulose synthase. The polynucleotide is
CC used for modulating, preferably increasing, the level of the synthase
CC in a plant cell. The plants are preferably monocots. The polynucleotide
CC is also used as a probe or primer in the detection quantitation or
CC isolation of gene transcripts. The probes are useful in detecting
CC deficiencies in the level of mRNA in screenings for desired transgenic
CC plant, for detecting mutations in the gene, for monitoring upregulation
CC of expression or changes in enzyme activity in screening assays of
CC compounds, for detection of any number of allelic variants of the gene,
CC or for use as molecular markers in plant breeding programs. The
CC isolated nucleic acids of the present invention can also be used for
CC recombinant expression of their encoded polypeptides or for use as
CC immunogens in the preparation and/or screening of antibodies. The
CC proteins can be employed in assays for enzyme agonists or antagonists
CC of enzyme function or for use of immunogens or antigens to obtain
CC antibodies specifically immunoreactive with a protein.

```
XX
SQ   Sequence 3725 BP; 927 A; 807 C; 950 G; 1041 T; 0 other;

Alignment Scores:
Pred. No.:                8.45e-31        Length:          3725
Score:                    40.00           Matches:         40
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              4.09%           Indels:          0
DB:                       21              Gaps:            0

US-09-980-043-2 (1-978) x AAZ99506 (1-3725)

Qy        256 PheValSerThrValAspProLeuLysGluProProLeuIleThrAlaAsnThrValLeu 275
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db       1232 TTTGTCAGTACAGTGGATCCATTGAAGGAACCTCCACTGATCACAGCCAACACTGTTTTG 1291

Qy        276 SerIleLeuAlaValAspTyrProValAspLysValSerCysTyrValSerAspAspGly 295
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1292 TCCATTCTTGCTGTGGATTACCCTGTTGACAAAGTGTCATGCTATGTTTCTGATGATGGC 1351

RESULT 12
AAZ99521
ID   AAZ99521 standard; DNA; 3725 BP.
XX
AC   AAZ99521;
XX
DT   03-JUL-2000  (first entry)
XX
DE   DNA encoding a maize cellulose synthase.
XX
KW   Maize; cellulose synthase; stalk quality; stand; silage; cellulose;
KW   transgenic plant; plant breeding marker; ss.
XX
OS   Zea mays
XX
FH   Key             Location/Qualifiers
FT   CDS             179..3397
FT                   /*tag= a
FT                   /product= "cellulose synthase"
FT                   /note= "no termination codon given"
XX
PN   WO200037706-A2.
XX
PD   29-JUN-2000
XX
PF   16-AUG-1999;   99WO-US18760.
XX
PR   17-AUG-1998;   98US-0096822.
XX
PA   (PION-) PIONEER HI-BRED INT INC.
XX
PI   Dhugga KS, Helentjaris TG, Bowen BA, Wang X.
XX
DR   WPI; 2000-224343/19.
DR   P-PSDB; AAY84117
XX
PT   New genes which encode maize cellulose synthase polypeptides in plants
PT   useful for modulating the expression of cellulose synthase in plants
PT   and to produce transgenic plants expressing the novel protein -
XX
PS   Claim 1; Page 160-165; 119pp; English.
XX
CC   The present sequence encodes a maize cellulose synthase polypeptide.
CC   The cellulose synthase can be used for the improvement of stalk quality
CC   for improved stand or silage. It also provides an increased concentration
CC   of cellulose in the pericarp, hardening the kernel and improving its
CC   handling ability. The sequences are used to produce transgenic plants
CC   and seeds expressing the cellulose synthase. The polynucleotide is
CC   used for modulating, preferably increasing, the level of the synthase
CC   in a plant cell. The plants are preferably monocots. The polynucleotide
CC   is also used as a probe or primer in the detection, quantitation or
CC   isolation of gene transcripts. The probes are used in detecting
CC   deficiencies in the level of mRNA in screenings for desired transgenic
CC   plant, for detecting mutations in the gene, for monitoring upregulation
CC   of expression or changes in enzyme activity in screening assays of
CC   compounds, for detection of any number of allelic variants of the gene,
CC   or for use as molecular markers in plant breeding programs. The
CC   isolated nucleic acids of the present invention can also be used for
CC   recombinant expression of their encoded polypeptides or for use as
CC   immunogens in the preparation and/or screening of antibodies. The
CC   proteins can be employed in assays for enzyme agonists or antagonists
CC   of enzyme function or for use of immunogens or antigens to obtain
CC   antibodies specifically immunoreactive with a protein.
XX
SQ   Sequence 3725 BP; 927 A; 807 C; 950 G; 1041 T; 0 other;

Alignment Scores:
Pred. No.:                8.45e-31        Length:          3725
Score:                    40.00           Matches:         40
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              4.09%           Indels:          0
DB:                       21              Gaps:            0

US-09-980-043-2 (1-978) x AAZ99521 (1-3725)

Qy        256 PheValSerThrValAspProLeuLysGluProProLeuIleThrAlaAsnThrValLeu 275
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db       1232 TTTGTCAGTACAGTGGATCCATTGAAGGAACCTCCACTGATCACAGCCAACACTGTTTTG 1291

Qy        276 SerIleLeuAlaValAspTyrProValAspLysValSerCysTyrValSerAspAspGly 295
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1292 TCCATTCTTGCTGTGGATTACCCTGTTGACAAAGTGTCATGCTATGTTTCTGATGATGGC 1351

RESULT 13
AAA67086
ID   AAA67086 standard; DNA; 1407 BP.
XX
AC   AAA67086;
XX
DT   31-OCT-2000  (first entry)
XX
DE   Eucalyptus grandis cellulose synthase nucleotide sequence SEQ ID NO:14.
XX
KW   Eucalyptus grandis; pinus radiata; Monterey pine; modification;
KW   plant cell wall; polysaccharide; polysaccharide biosynthetic pathway;
KW   transgenic plant; ds.
XX
OS   Eucalyptus grandis.
XX
PN   WO200022092-A2.
XX
PD   20-APR-2000
XX
PF   08-OCT-1999;   99WO-NZ00169.
XX
PR   13-OCT-1998;   98US-0170862.
PR   11-AUG-1999;   99US-0148426.
XX
PA   (GENE-) GENESIS RES & DEV CORP LTD.
PA   (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI   Bloksberg LN;
XX
DR   WPI; 2000-339328/29.
DR   P-PSDB; AAB16281.
XX
PT   New genes encoding proteins involved in a plant polysaccharide
PT   biosynthetic pathway, useful for modulating or altering the
PT   polysaccharide content, composition or structure of the plant -
XX
PS   Claim 1; Page 42-43; 301pp; English.
XX
CC   The present invention describes isolated polynucleotides (PN) comprising
CC   a sequence selected from one of 835 nucleotide sequences given in
```

AAA67073 to AAA67907, their (reverse) complements, sequences producing an Expectation (E) value of 0.01 or less compared to the 835 sequences, sequences at least 50% identical to them, 200, 100, 40 or 20-mers of the 835 sequences or sequences that are degenerately equivalent or allelic to the 835 sequences. The polynucleotides are used to modify the activity of a polypeptide involved in a polysaccharide biosynthetic pathway in the plant. They are especially used to modulate or alter the polysaccharide content, composition or structure of the plant. AAB16268 to AAB16340 are proteins encoded by some of the polynucleotide sequence given in the present invention.

Sequence 1407 BP; 365 A; 296 C; 355 G; 391 T; 0 other;

ignment Scores:
ed. No.: 3.68e-30 Length: 1407
ore: 39.00 Matches: 39
rcent Similarity: 100.00% Conservative: 0
st Local Similarity: 100.00% Mismatches: 0
ery Match: 3.99% Indels: 0
21 Gaps: 0

-09-980-043-2 (1-978) x AAA67086 (1-1407)

512 ValPhePheAspValAsnMetLysGlyLeuAspGlyIleGlnGlyProValTyrValGly 531
||||||||||||||||||||||||||||||||||||||||||||||||||||
715 GTTTTCTTTGATGTTAACATGAAAGGGCTGGATGGAATCCAAGGACCAGTTTATGTCGGG 774

532 ThrGlyCysValPheAsnArgGlnAlaLeuTyrGlyTyrGlyProProSerMetPro 550
|||||||||||||||||||||||||||||||||||||||||||||||||||||
775 ACAGGATGTGTTTTCAATAGGCAAGCACTTTACGGATACGGACCCCCTTCTATGCCG 831

SULT 14
V34433
AAV34433 standard; cDNA; 2306 BP.

AAV34433;

28-SEP 1998 (first entry)

Cotton cellulose synthase CelA2 partial cDNA.

Cellulose synthase; CelA2 gene; cotton; forest tree; plant fibre; wood; transgenic plant; ds.

Gossypium hirsutum Acala SJ-2.

Key Location/Qualifiers
CDS 39..2099
/*tag= a

WO9818949-A2.

07-MAY-1998.

29-OCT-1997; 97WO-US19529.

29-OCT-1996; 96US-0029987.

(CALJ) CALGENE INC.

Delmer DP, Pear JR, Stalker DM;

WPI; 1998-272238/24.
P-PSDB; AAW60040.

New isolated plant cellulose synthesis enzyme DNA - is used to develop products for producing modified plant fibre products, particularly for modifying cotton fibre and wood quality

Claim 7; Fig 7A-D; 49pp; English.

This partial cDNA sequence includes approximately the last 3' two-thirds of the cotton CelA2 gene transcript, and encodes a

CC 686-amino acid cellulose synthase (see AAW60040). 2 cDNA clones
CC that exhibited small blocks of amino acid homology to proteins
CC encoded by bacterial CelA genes were isolated from an unamplified
CC cDNA library prepared from mRNA of fibres of cotton Acala SJ-2
CC harvested at 21 dpa (stage of maximal secondary wall cellulose
CC synthesis). These 2 clones were partially homologous at the
CC nucleotide and amino acid levels and were designated CelA1 (see
CC AAV34432) and CelA2, respectively. The 2 cotton CelA genes are
CC both expressed in developing fibre at approximately 500 times their
CC level of expression in other cotton tissues, and constitute
CC approximately 1.2% of the 24 dpa fibre mRNA. The CelA1 gene
CC promoter region (see AAV34434) is also claimed. The isolated cDNA
CC clones are used in cla[illegible] methods for modifying fibre phenotype
CC in a cotton plant and for modifying wood quality phenotype in a
CC forest tree species. [illegible] can be reduced by over-expressing the
CC cellulose synthase [illegible] and disrupting production into cellulose.
XX
SQ Sequence 2306 BP; [illegible] other;

Alignment Scores:
Pred. No.: [illegible] Length: 2306
Score: 39.00 Matches: 39
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.99% Indels: 0
DB: 19 Gaps: 0

US-09-980-043-2 (1-978) x AAV34433 (1-2306)

Qy 417 GluLeuProArgGlnValTyrValSerArgGlnLysArgProGlyTyrGlnHisHisLys 436
||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 354 GAGCTGCCTCGACAGGTCTATGTTTCTCGTCAGAAACGACCTGGTTATCAGCACCATAAG 413

Qy 437 LysAlaGlyAlaGlyAsnAlaLeuValArgValSerAlaValLeuThrAsnAlaPro 455
||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 414 AAAGCCGGTGCTGGAAATGCTCTGGTTCGAGTTTCTGCAGTGCTTACTAATGCACCC 470

RESULT 15
AAV08373
ID AAV08373 standard; cDNA to mRNA; 3111 BP.
XX
AC AAV08373;
XX
DT 08-FEB-1999 (first entry)
XX
DE Cellulose synthase PesA coding sequence.
XX
KW Cellulose synthase; cotton; ds.
XX
OS Gossypium hirsutum.
XX
FH Key Location/Qualifiers
FT CDS 22..3114
FT /*tag= a
XX
PN EP875575-A2.
XX
PD 04-NOV-1998.
XX
PF 31-MAR-1998; 98EP-0302489.
XX
PR 01-APR-1997; 97JP-0081133.
XX
PA (NISN) NISSHINBO IND INC.
XX
PI Aotsuka S, Hasegawa O, Hayashi T, Ihara Y;
XX
DR WPI; 1998-559440/48.
DR P-PSDB; AAW73309.
XX
PT New DNA encoding cotton cellulose synthase proteins - useful for
PT controlling cellulose synthesis in prokaryotic or eukaryotic cells
XX

PS Example 5; Page 22-27; 51pp; English.
XX
CC This sequence encodes a cellulose synthase of the invention. The DNA can
CC be used for controlling cellulose synthesis in a cell by introducing the
CC DNA into the cell, expressing RNA with a nucleotide sequence homologous
CC to the DNA or a nucleotide sequence complementary to the DNA. The cells
CC are prokaryotic cells i.e. acetobacterium and/or eukaryotic cells such as
CC yeasts i.e. Saccharomyces spp., plant cells (cotton plants) and cultured
CC cells of mammals etc.
XX
SQ Sequence 3311 BP; 900 A; 678 C; 798 G; 945 T; 0 other;

Alignment Scores:

Pred. No.:	8.29e-30	Length:	3311
Score:	39.00	Matches:	20
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	3.99%	Indels:	0
DB:	19	Gaps:	0

US-09-980-043-2 (1-978) x AAV08373 (1-3311)

Qy 417 GluLeuProArgLeuValTyrValSerArgGluLysArgProGlyTyrGlnHisHisLys 436
Db 1397 GAGCTGCCTCGACTTGTCTATGTTTCTCGTGAGAAACGACCTGGTTATCAGCACCATAAG 1456

Qy 437 LysAlaGlyAlaGluAsnAlaLeuValArgValSerAlaValLeuThrAsnAlaPro 455
Db 1457 AAAGCCGGTGCTGAGAATGCTCTTGTTCGAGTTTCTGCAGTTTTACTAATGCACCC 1513

Search completed: October 31, 2003, 23:55:45
Job time : 746.785 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 31, 2003, 20:56:23 ; Search Time 10527.2 Seconds
(without alignments)
3403.589 Million cell updates/sec

Title: US-09-980-043-2
Perfect score: 978
Sequence: 1 MMESGAPICHTCGEQVGHDA.....FFVNKVNSLVAETCISIDC 978

Scoring table: OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 2045481998 residues

Word size: 1

Total number of hits satisfying chosen parameters: 5773148

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post processing: Listing first 45 summaries

Command line parameters:
MODEL=frame_p2n.model -DEV=xlb
-Q /cgn2_1/USPTO_spool/US09980043/runat_29102003_16394_22663/app_query.fasta_1.1358
DB=GenEmbl QFMT=fastap -SUFFIX=oligo.rge MINMATCH=0 -LOOPCUT=0 LOOPEXT=0
UNITS=bits -START=1 -END=-1 -MATRIX=oligo TRANS human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1 ALIGN 15 MODE=LOCAL -OUTFMT=pto
NORM ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN 2000000000
USER=US09980043_aCGN_1_1_6944_srunat_29102003_163930 ..963 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 WAIT DSPBLOCK=100 LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 XGAPOP=60 XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 YGAPOP=60 -YGAPEXT=60 -DELOP=6 DELEXT=7

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		978	100.0	3232	8	AF072131	AF072131 Populus t
2		92	9.4	3145	8	AF323039	AF323039 Zinnia el
3		61	6.2	3186	8	GHU58283	U58283 Gossypium h
4		61	6.2	3328	6	AR164434	AR164434 Sequence
5		61	6.2	6886	8	AF413210	AF413210 Gossypium
6		47	4.8	1515	8	AF323040	AF323040 Zinnia el
7		47	4.8	1518	8	AF323041	AF323041 Zinnia el
8		47	4.8	2958	8	AF267742	AF267742 Arabidops
9		47	4.8	3109	8	AY099636	AY099636 Arabidops
10		44	4.5	1010	8	AF197911	AF197911 Populus t
11	c	43	4.4	94301	8	ATF28A21	AL035526 Arabidops
12	c	43	4.4	199075	8	ATCHRIV49	AL161549 Arabidops
13		40	4.1	1741	6	AR267561	AR267561 Sequence
14		40	4.1	1741	6	AX030950	AX030950 Sequence
15		40	4.1	1741	6	BD022680	BD022680 Manipulat
16		40	4.1	1838	8	AF030052	AF030052 Oryza sat
17		40	4.1	3725	8	AF200526	AF200526 Zea mays
18	c	40	4.1	50149	2	AC132494	AC132494 Oryza sat
19	c	40	4.1	142160	2	AC135426	AC135426 Oryza sat
20	c	40	4.1	157010	2	AC144738	AC144738 Oryza sat
21		39	4.0	2229	8	GHU58284	U58284 Gossypium h
22		39	4.0	3168	8	AF458083	AF458083 Arabidops
23		39	4.0	3181	8	BT006111	BT006111 Arabidops
24		39	4.0	3377	8	BT005710	BT005710 Arabidops
25		39	4.0	71522	8	AB006703	AB006703 Arabidops
26		35	3.6	3229	8	BT002335	BT002335 Arabidops
27		35	3.6	3614	6	AR267559	AR267559 Sequence
28		35	3.6	3614	6	AX030946	AX030946 Sequence
29		35	3.6	3614	6	BD022678	BD022678 Manipulat
30	c	35	3.6	29292	8	AB018111	AB018111 Arabidops
31		32	3.3	3255	6	AX505864	AX505864 Sequence
32		32	3.3	3255	6	AX507835	AX507835 Sequence
33		32	3.3	3444	8	AF062485	AF062485 Arabidops
34		32	3.3	3771	8	AF027173	AF027173 Arabidops
35		32	3.3	3828	6	AR267558	AR267558 Sequence
36		32	3.3	3828	6	AX030944	AX030944 Sequence
37		32	3.3	3828	6	BD022677	BD022677 Manipulat
38		32	3.3	5009	6	AR267556	AR267556 Sequence
39		32	3.3	5009	6	AX030941	AX030941 Sequence
40		32	3.3	5009	6	BD022675	BD022675 Manipulat
41		32	3.3	39645	8	AB025637	AB025637 Arabidops
42		32	3.3	89035	8	ATT22F8	AL050351 Arabidops
43	c	32	3.3	116428	2	AC135232	AC135232 Medicago
44	c	32	3.3	119004	2	AC137522	AC137522 Medicago
45		32	3.3	162776	8	AP003237	AP003237 Oryza sat

ALIGNMENTS

RESULT 1

```
72131
US        AF072131                 3232 bp    mRNA    linear   PLN 26 APR-2001
INITION   Populus tremuloides secondary xylem cellulose synthase (celA) mRNA,
          complete cds.
ESSION    AF072131
SION      AF072131.1  GI:4115904
WORDS     .
RCE       Populus tremuloides (quaking aspen)
RGANISM   Populus tremuloides
          Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
          Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
          rosids; eurosids I; Malpighiales; Salicaceae; Populus.
ERENCE    1  (bases 1 to 3232)
UTHORS    Wu,L., Joshi,C.P. and Chiang,V.L.
ITLE      A xylem-specific cellulose synthase gene from aspen (Populus
          tremuloides) is responsive to mechanical stress
OURNAL    Plant J. 22 (6), 495-502 (2000)
EDLINE    20344632
PUBMED    10886769
ERENCE    2  (bases 1 to 3232)
UTHORS    Wu,L., Joshi,C.P. and Chiang,V.L.
ITLE      Direct Submission
OURNAL    Submitted (14-JUN-1998) IWR/Forestry, Michigan Tech University,
          1400, Townsend Drive, Houghton, MI 49931, USA
ERENCE    3  (bases 1 to 3232)
UTHORS    Wu,L., Joshi,C.P. and Chiang,V.L.
ITLE      Direct Submission
OURNAL    Submitted (07-JAN-1999) IWR/Forestry, Michigan Tech University,
          1400, Townsend Drive, Houghton, MI 49931, USA
EMARK     Sequence update by submitter
MENT      On Jan 8, 1999 this sequence version replaced gi:3243277.
ATURES             Location/Qualifiers
     source        1..3232
                   /organism="Populus tremuloides"
                   /mol_type="mRNA"
                   /db_xref="taxon:3693"
                   /tissue_type="secondary xylem"
     gene          1..3232
                   /gene="celA"
     CDS           69..3005
                   /gene="celA"
                   /codon_start=1
                   /product="secondary xylem cellulose synthase"
                   /protein_id="AAD03417.1"
                   /db_xref="GI:4115905"
                   /translation="MMESGAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEI
                   NEGRKVCLRCGSPYDENLLDDVEKKGSGNQSTMASHLNDSQDVGIHARHISSVSTVDS
                   EMNDEYGNPIWKNRVKSCKDKENKKKKRSPKAETEPAQVPTEQQMEEKPSAEASEPLS
                   IVYPIPRNKLTPYRAVIIMRLVILGLFFHFRITNPVDSAFGLWLTSVICEIWFAFSWV
                   LDQFPKWNPVNRETYIERLSARYEREGEPSQLAGVDFFVSTVDPLKEPPLITANTVLS
                   ILAVDYPVDKVSCYVSDDGAAMLSFESLVETAEFARKWVPFCKKFSIEPRAPEFYFSQ
                   KIDYLKDKVQPSFVKERRAMKRDYEEYKVRVNALVAKAQKTPEEGWTMQDGTPWPGNN
                   TRDHPGHDSGLPWEILGARDIEGNELPRLVYVSREKRPGYQHHKKAGAENALVRVSAV
                   LTNAPYILNVDCDHYVNNSKAVREAMCILMDPQVGRDVCYVQFPQRFDGIDKSDRYAN
                   RNVVFFDVNMKGLDGIQGPVYVGTGCVFNRQALYGYGPPSMPSLRKRKDSSSCFSCCC
                   PSKKKPAQDDPAEVYRDAKREDLNAAIFNLTEIDNYDEHERSMLISQLSFEKTFGLSSV
                   FIESTLMENGGVPESANSPPFIKEAIQVIGCGYEEKTEWGKQIGWIYGSVTEDILSGF
                   KMHCRGWRSIYCMPVRPAFKGSAPINLSDRLHQVLRWALGSVEIFFSRHCPLWYGFGG
                   GRLKWLQRLAYINTIVYPFTSLPLIAYCTIPAVCLLTGKFIIPTLSNLASMLFLGLFI
                   SIIVTAVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQGFLKMLAGIDTNFTVTA
                   KAAEDAEFGELYMVKWTTLLIPPTTLLIINMSGCAGFSDALNKGYEAWGPLFGKVFFA
                   FWVILHLYPFLKGLMGRQNLTPTIVVLWSVLLASVFSLVWVKINPFVNKVDNTLVAET
                   CISIDC"
SE COUNT      895 a    668 c    772 g    897 t
IGIN

ignment Scores:
ed. No.:              0              Length:        3232
ore:                  978.00         Matches:       978
rcent Similarity:     100.00%        Conservative:  0
st Local Similarity:  100.00%        Mismatches:    0
ery Match:            100.00%        Indels:        0
                      8              Gaps:          0

US-09-980-043 2 (1 97[illegible] x A[illegible] 1 [illegible]

Qy          1 MetMetGluSerGl[illegible]CysHisCysThrCysGlyGluGlnValGlyHisAspAla 20
              ||| ||| ||| ||| ||  ||| |||| |||| |||| |||| |||| ||
Db         69 ATGATGGAATCTGG[illegible]TGTGGTGAACAGGTGGGGCATGATGCA 120

Qy         21 AsnGlyGluGlu[illegible]CysHisSerCysTyrProMetCysLysSerCysPhe 40
              |||| ||| ||  ||| ||  ||| |||| |||| |||| |||| |||| |
Db        129 AATGGGGAGGAT[illegible]GCTATGCCATGTGCAAGTCTTGTTTC 180

Qy         41 GluPheGlu[illegible]LysTyrLysValCysLeuArgCysGlySerProTyrAspGlu 60
              |||| |||  ||| ||| ||| |||| |||| |||| |||| |||| |
Db        189 GAGTTTGAAA[illegible]AAGTTTGCGGTGTGGCTCGCCATATGATGAG 240

Qy         61 AsnLeuLeuAsp[illegible]GlyAsnGlnSerThrMetAlaSerHis 80
              ||| ||| |||  ||| ||| ||| |||| |||| |||| |||| |||| |
Db        249 AACTTGCTAGA[illegible]AATCAATCGATAATCATCTTCAC 294

Qy         81 LeuAsnA[illegible]HisIleSerSerValSerThrVal 100
              ||| || |||  ||| ||| ||| |||| |||| |||| |||| |
Db        309 CTCAAGGATT[illegible]AGACATATCAGTAGTGTGTCCACTGTG 360

Qy        101 AspSerGluMetAs[illegible]GlyAsnProIleTrpLysAsnArgValLysCys 120
              ||| ||| |||  ||| ||| ||| |||| |||| |||| |||| |||| |
Db        369 GATAGTGAAATGAA[illegible]AATTTGGAAGAATCGGGTGAAGAGTGT 420

Qy        121 LysAspLys[illegible]AsnLysLysLysArg[illegible]ProLysAlaGluThrGluProAlaGln 140
              ||| ||| |||  ||| ||| ||| |||| |||| |||| |||| |||| |
Db        429 AAGGATAAA[illegible]AAAAAAAGGAA[illegible]TAAGGCTGAAACTGAACCAGCTCAA 480

Qy        141 ValProThrGluGl[illegible]MetGlyGluGlyProSerAlaGluAlaSerGluProLeuSer 160
              |||| |||  ||| ||| ||| |||| |||| |||| |||| |||| |
Db        489 GTTCCTACAGAAGA[illegible]AGATGGAAGAAAAAGTCTTGCAGAGGCTTCGGAGCCGCTTTCA 540

Qy        161 IleValTyrProIl[illegible]ProArgAsnLysLeuThrProTyrArgIleValIleIleMetArg 180
              |||| |||  ||| ||| ||| |||| |||| |||| |||| |||| |
Db        549 ATTGTTTATCCAAT[illegible]ACGCAAGAAATCTACACTATACAGAGCAGTGATCATTATGCGA 600

Qy        181 LeuValIleLeuGl[illegible]PheHisPheArgIleThrAsnProValAspSerAlaIle 200
              ||| ||| |||  ||| |||| |||| |||| |||| |||| |||| |
Db        609 CTGGTCATTCTGGG[illegible]TTCTTCCACTTCAGAATAACAAATCCTGTCGATAGTGCCTTT 660

Qy        201 GlyLeuTrpLeuThr[illegible]ValIleCysIleIleThrPheIleAlaPheSerTrpValLeuAsp 220
              ||| ||| |||| |||| ||| |||| |||| |||| |||| |||| |
Db        669 GGCCTGTGGCTTAC[illegible]TTGTGATATTTGGAGACTTGGTTTGCATTTTCTTGGGTGTTGGAT 720

Qy        221 GlnPheProLysTr[illegible]AsnProValAsnArgGluThrTyrIleGluArgLeuSerAlaArg 240
              ||| ||| |||  ||| ||| ||| |||| |||| |||| |||| |||| |
Db        729 CAGTTCCCCAAGTG[illegible]AATCCTGTCAATAGAGAAACGTATATCGAAAGGCTGTCGGCAAGG 780

Qy        241 TyrGluArgGluGl[illegible]ProSerGlnLeuAlaGlyValAspPhePheValSerThrVal 260
              ||| ||| |||  ||| ||| ||| |||| |||| |||| |||| |||| |
Db        789 TATGAAAGAGAGGT[illegible]TCCTCTCAGCTCGCTGGTGTGGATTTTTTCGTGAGTACTGTT 840

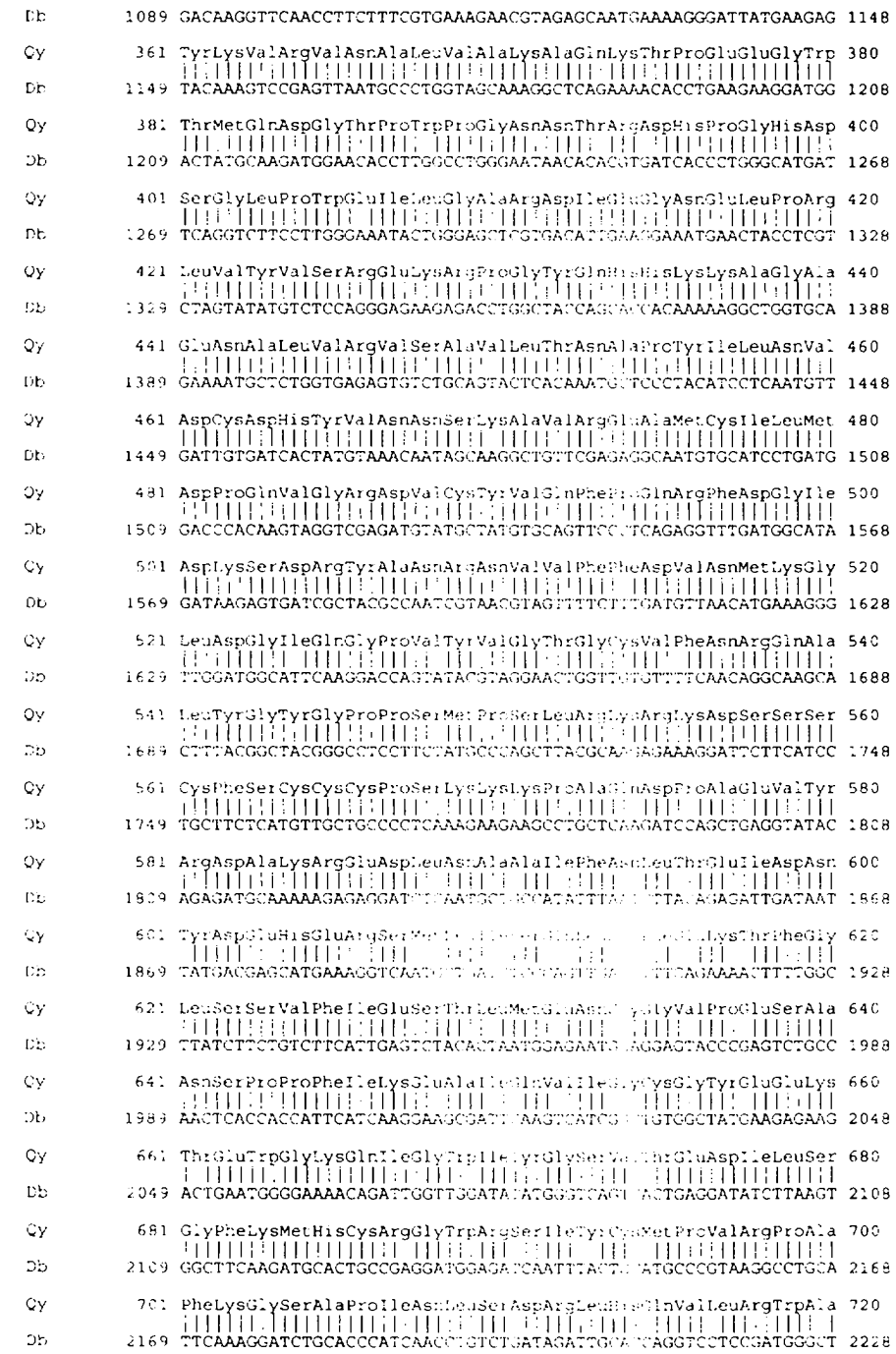
Qy        261 AspProLeuLysGl[illegible]ProLeuIleThrAlaAsnThrValLeuSerIleLeuAlaVal 280
              ||| ||| |||  ||| ||| ||| |||| |||| |||| |||| |||| |
Db        849 GATCCGCTGAAGGA[illegible]TGTCATTGACTGCAATACAGTCCTTTCCATCCTTGCTGTG 900

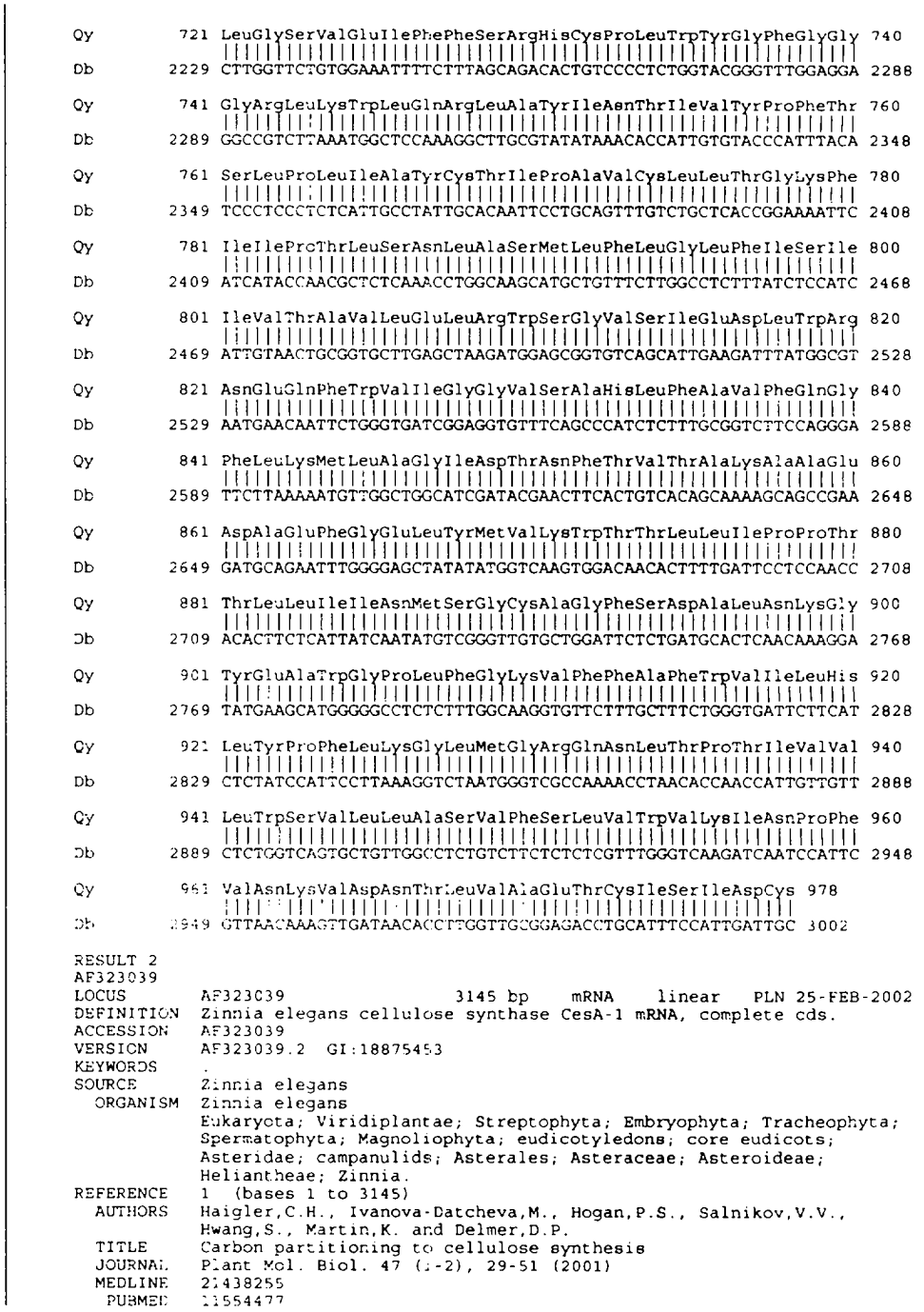
Qy        281 AspTyrProValAr[illegible]LysValSerLysTyrValSerAspAspGlyAlaAlaMetLeuSer 300
              ||| ||| |||  ||| ||| ||| |||| |||| |||| |||| |||| |
Db        909 GACTATCCCGTCGA[illegible]AAGTTTCCTGCTATGTCTCTGATGATGGTGCAGCTATGCTTTCA 960

Qy        301 PheGluSerLeuVa[illegible]GluThrAlaGluPheAlaArgLysTrpValProPheCysLysLys 320
              |||| ||| |||  ||| ||| ||| |||| |||| |||| |||| |||| |
Db        969 TTTGAATCTCTTGTA[illegible]AAACAGCTCAGTTTGCAAGGAAGTGGGTTCCGTTCTGCAAAAAA 1020

Qy        321 PheSerIleGluPr[illegible]ArgAlaProGluPheTyrPheSerGlnLysIleAspTyrLeuLys 340
              ||| ||| |||  ||| ||| ||| |||| |||| |||| |||| |||| |
Db       1029 TTCTCAATTGAACC[illegible]AAGCACCGGAGTTTTACTTCTCACAGAAAATTGATTACTTGAAA 1080

Qy        341 AspLysValGlnProSerPheValLysGluArgArgAlaMetLysArgAspTyrGluGlu 360
              ||| ||| |||  ||| ||| ||| |||| |||| |||| |||| |||| ||
```





```
RESULT 2
AF323039
LOCUS       AF323039               3145 bp    mRNA    linear   PLN 25-FEB-2002
DEFINITION  Zinnia elegans cellulose synthase CesA-1 mRNA, complete cds.
ACCESSION   AF323039
VERSION     AF323039.2  GI:18875453
KEYWORDS    .
SOURCE      Zinnia elegans
  ORGANISM  Zinnia elegans
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; campanulids; Asterales; Asteraceae; Asteroideae;
            Heliantheae; Zinnia.
REFERENCE   1  (bases 1 to 3145)
  AUTHORS   Haigler,C.H., Ivanova-Datcheva,M., Hogan,P.S., Salnikov,V.V.,
            Hwang,S., Martin,K. and Delmer,D.P.
  TITLE     Carbon partitioning to cellulose synthesis
  JOURNAL   Plant Mol. Biol. 47 (1-2), 29-51 (2001)
  MEDLINE   21438255
   PUBMED   11554477
```

```
 RENCE   2  (bases 1 to 3145)
 THORS   Hwang,S. and Haigler,C.H.
  TLE    Direct Submission
 URNAL   Submitted (22-NOV-2000) Biological Sciences, Texas Tech University,
         Box 43131, Lubbock, TX 79409, USA
 RENCE   3  (bases 1 to 3145)
 THORS   Hwang,S. and Haigler,C.H.
  TLE    Direct Submission
 URNAL   Submitted (25-FEB-2002) Biological Sciences, Texas Tech University,
         Box 43131, Lubbock, TX 79409, USA
 MARK    Sequence update by submitter
 MENT    On Feb 25, 2002 this sequence version replaced gi:13021930.
 URES             Location/Qualifiers
   source         1..3145
                  /organism="Zinnia elegans"
                  /mol_type="mRNA"
                  /db_xref="taxon:34245"
   CDS            69..3005
                  /codon_start=1
                  /product="cellulose synthase CesA-1"
                  /protein_id="AAK11588.2"
                  /db_xref="GI:18875454"
                  /translation="MMHPETPICNTCGEHIGFGPDGNGVFVGCHECNFPICKSCLEYE
                  INEGHTACIQCGTPYDGNLTKVGEMEKEPVTRNTMAAHLSNSQDAGLHARNVSTVSTV
                  DTEMTDESGNPIWKNRVESWKDKKNKKKKGGDKVAKEVQVPEDQHIEEKQQSADPNAM
                  QPPSQIIPIPKSQITPYRIVIILRLIILGLFFHYRITNPVESSYGLWLTSVICEIWFA
                  ISWVLDQFPKWYPINRVTFTDELSARYEREGEPSELAAVDFFVSTVDPLKEPPLITAN
                  TVLSILAVDYPVEKVSCYVSDDGAAMLTFESLVETAEFARKWVPFCKKFSIEPRAPEF
                  YFSQKIDYLKDKVQPSFVKERRAMKRDYEEYKVRVNALVAKAQKTPEEGWTMQDGTPW
                  PGNNPRDHPGMIQVFLGHSGAHDIEGNELPRLVYVSREKRPGYQHHKKAGAENALVRV
                  SAVLTNAPYILNLDCDHYVNNSKAVREAMCFMMDPQVGRDVCYIQFPQRFDGIDRSDR
                  YANRNTVFFDVNMKGLDGIQGPVYVGTGCVFYRQALYGYGPQSLPTLPSPSSSSSCCC
                  CGPKKPKKDLEEFKRDARRDDLNAAIFNLKEIESYDDYERSLLISQMSFEKTFGMSSV
                  FIESTLMENGGLAESANPATMINEAIHVISCGYEEKTAWGKEIGWIYGSVTEDILTGF
                  KMHCRGWRSIYCMPVRPAFKGSAPINLSDRLHQVLRWALGSVEIFLSRHCPLWYGWGG
                  GRLKLLQRLAYINTIVYPFTSLPLVAYCTLPAICLLTGKFIIPTLSNIAAVWFLGLFL
                  SIITTSVLEIRWSGVSIEELWRNEQFWVIGGVSAHLFAVFQGFLKMLAGVDTNFTVTA
                  KAADDQEFGELYMIKWTTVLIPPTTLLVLNLVGVVAGFSDALNKGYEAWGPLFGKVFF
                  AFWVILHLYPFLKGLMGRQNRTPTIVILWSVLLASVFSLVWVKIDPFVSKGDSNLTQG
                  CIAIDC"
 E COUNT      873 a    582 c    746 g    944 t
 GIN

 ignment Scores:
 ed. No.:              3.36e-92       Length:        3145
 ore:                  92.00          Matches:       92
 rcent Similarity:     100.00%        Conservative:  0
 st Local Similarity:  100.00%        Mismatches:    0
 ery Match:            9.41%          Indels:        0
                       8              Gaps:          0

-09-980-043-2 (1-978) x AF323039 (1-3145)

       301 PheGluSerLeuValGluThrAlaGluPheAlaArgLysTrpValProPheCysLysLys 320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
       981 TTTGAATCTTTAGTTGAAACAGCAGAGTTTGCAAGAAAGTGGGTCCCATTTTGTAAAAAA 1040

       321 PheSerIleGluProArgAlaProGluPheTyrPheSerGlnLysIleAspTyrLeuLys 340
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
      1041 TTCTCGATAGAACCACGCGCACCTGAGTTTTACTTCTCTCAGAAGATTGACTACTTGAAA 1100

       341 AspLysValGlnProSerPheValLysGluArgArgAlaMetLysArgAspTyrGluGlu 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
      1101 GATAAAGTGCAGCCATCGTTTGTCAAAGAACGTAGAGCAATGAAAAGGGATTACGAAGAG 1160

       361 TyrLysValArgValAsnAlaLeuValAlaLysAlaGlnLysThrProGluGluGlyTrp 380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
      1161 TACAAAGTGCGGGTCAATGCTTTAGTAGCAAAAGCTCAAAAAACTCCTGAAGAAGGATGG 1220

       381 ThrMetGlnAspGlyThrProTrpProGlyAsnAsn 392
           ||||||||||||||||||||||||||||||||||||
      1221 ACTATGCAAGATGGAACACCTTGGCCTGGAAATAAC 1256
```

```
RESULT 3
GHU58283
LOCUS       GHU58283                3186 bp    mRNA    linear   PLN 05-DEC-1996
DEFINITION  Gossypium hirsut[illegible] cellulose synthase (celA1) mRNA, complete cds.
ACCESSION   U58283
VERSION     U58283.1  GI:17[illegible]
KEYWORDS    .
SOURCE      Gossypium hirsutum (upland cotton)
  ORGANISM  Gossypium hirsutum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; M[illegible]oliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium.
REFERENCE   1  (bases 1 to [illegible])
  AUTHORS   Pear,J.R., Kawa[illegible],Y., Schr[illegible],W.E., Delmer,D.P. and
            Stalker,D.M.
  TITLE     Higher plants [illegible] the bacterial celA genes encoding
            the catalyt[illegible] synthase
  JOURNAL   Proc. Natl. A[illegible] 12640-12643 (1996)
  MEDLINE   97057296
   PUBMED   8901639
REFERENCE   2  (bases 1 to [illegible])
  AUTHORS   Pear,J.R., Kawa[illegible],Y., Schr[illegible],W.E., Delmer,D.P. and
            Stalker,D.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-MAY 1996) J.R. Pear, Calgene, Inc, 1920 5th Street,
            Davis, CA 95616, USA
FEATURES             Location/Qualifiers
     source          1..3186
                     /organism="Gossypium hirsutum"
                     /mol_type="mRNA"
                     /strain="Acala SJ-2"
                     /db_xref="taxon:3635"
                     /tissue_type="fiber"
                     /dev_stage="17-21 days post anthesis"
     gene            1..3186
                     /gene="celA1"
     CDS             77..3[illegible]
                     /gene="celA1"
                     /codon_start=1
                     /product="cellulose synthase"
                     /protein_id="AAB37766.1"
                     /db_xref="GI:1706956"
                     /translation="MMESGVPVPHTCGEHVGLNVNGEPFVACHECNFPICKSCFEYDL
                     KEGRKACLRCGSPYDENLLDDVEKATGDQSTMAAHLNKSQDVGIHARHISSVSTLDSE
                     MAEDNGNSIWKNRVESWKEKKNKKKKPATTKVERRAEIPPEQQMEDKPAPDASQPLSI
                     IIPIPKSPLAPYRTVIIMRLIILGLFFHYRVTNPVDSAFGLWLTSVICEIWFAFSWVL
                     DQFPKWYPVNRETYIDRLSARYEREGEPDELAAVDFFVSTVDPLKEPPLITANTVLSI
                     LALDYPVEKVSCYISDDGAAMLTFESLVETADFARKWVPFCKKFSIEPRAPEFYFSQK
                     IDYLKDKVQPSFVKERRAMKRDYEEYKIRINALVAKAQKTPDEGWTMQDGTSWPGNNP
                     PDHPGMIQVFLGYSGARDIEGNELPRLVYVSREKRPGYQHHKKAGAENALVRVSAVLT
                     NAPFILNLDCDHYVNNSKAVREAMCFLMDPQVGRDVCYVQFPQRFDGIDRSDRYANRN
                     TVFFDVNMKGLDGIQGPVYVGTGCVFNRQALYGYGPPSMPSFPKSSSSSCSCCCPGKK
                     RPKDPSFVPPDAKPEELDAAIFNLREIDNYDEYERSMLISQTSFEKTFGLSSVFIESS
                     LMENGGVAESANPSTLIKEAIHVISCGYEEKTAWGKEIGWIYGSVTEDILTGFKMHCR
                     GWRSIYCMPQRPAFKGSAPINLSDRLHQVLRWALGSVEIFLSRHCPLWYGFGGGRLKW
                     LQRLAYINTIVYPFTSLPLIAYSTLPAICLLTGKFIIPTLSNLASVLFLGLFLSIIVT
                     AVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQGFLKMLAGIDTNFTVTAKAADD
                     ADFGELYMVKWTTLLIPPTTLLILNMVGVVAGFSDALNKGYEAWGPLFGKVFFSFWVI
                     LHLYPFLKGLMGRQNRTPTIVVLWSVLLASVFSLVWVRINPFVSTADSTTVSQSCISK
                     DC"
BASE COUNT      864 a    654 c    723 g    945 t
ORIGIN

Alignment Scores:
Pred. No.:              2.22e-57       Length:        3186
Score:                  61.00          Matches:       61
Percent Similarity:     100.00%        Conservative:  0
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            6.24%          Indels:        0
DB:                     8              Gaps:          0

US-09-980-043-2 (1-978) x GHU58283 (1-3186)
```

```
Qy        799 SerIleIleValThrAlaValLeuGluLeuArgTrpSerGlyValSerIleGluAspLeu 818
              ||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db       2456 TCCATTATCGTGACTGCTGTTCTCGAGCTCCGATGGAGTGGTGTCAGCATTGAGGACTTA 2515

Qy        819 TrpArgAsnGluGlnPheTrpValIleGlyGlyValSerAlaHisLeuPheAlaValPhe 838
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2516 TGGCGTAACGAGCAGTTTTGGGTCATCGGTGGCGTTTCAGCGCATCTCTTTGCCGTCTTC 2575

Qy        839 GlnGlyPheLeuLysMetLeuAlaGlyIleAspThrAsnPheThrValThrAlaLysAla 858
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2576 CAAGGTTTCCTTAAGATGCTTGCGGGCATTGACACCAACTTTACTGTCACTGCCAAAGCA 2635

Qy        859 Ala 859
              |||
Db       2636 GCT 2638

RESULT 4
AR164434
LOCUS       AR164434                3328 bp    DNA     linear   PAT 17-OCT-2001
DEFINITION  Sequence 1 from patent US 6271443.
ACCESSION   AR164434
VERSION     AR164434.1  GI:16235654
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 3328)
  AUTHORS   Stalker,D.M., Pear,J.R. and Delmer,D.
  TITLE     Cotton and rice cellulose synthase DNA sequences
  JOURNAL   Patent: US 6271443-A 1 07-AUG-2001;
FEATURES             Location/Qualifiers
     source          1..3328
                     /organism="unknown"
BASE COUNT      914 a    689 c    764 g    962 t
ORIGIN

Alignment Scores:
Pred. No.:                2.3e-57         Length:          3328
Score:                    61.00           Matches:         61
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              6.24%           Indels:          0
DB:                       6               Gaps:            0

US-09-980-043-2 (1-978) x AR164434 (1-3328)

Qy        799 SerIleIleValThrAlaValLeuGluLeuArgTrpSerGlyValSerIleGluAspLeu 818
              ||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db       2568 TCCATTATCGTGACTGCTGTTCTCGAGCTCCGATGGAGTGGTGTCAGCATTGAGGACTTA 2627

Qy        819 TrpArgAsnGluGlnPheTrpValIleGlyGlyValSerAlaHisLeuPheAlaValPhe 838
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2628 TGGCGTAACGAGCAGTTTTGGGTCATCGGTGGCGTTTCAGCGCATCTCTTTGCCGTCTTC 2687

Qy        839 GlnGlyPheLeuLysMetLeuAlaGlyIleAspThrAsnPheThrValThrAlaLysAla 858
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2688 CAAGGTTTCCTTAAGATGCTTGCGGGCATTGACACCAACTTTACTGTCACTGCCAAAGCA 2747

Qy        859 Ala 859
              |||
Db       2748 GCT 2750

RESULT 5
AF413210
LOCUS       AF413210                6886 bp    DNA     linear   PLN 27-DEC-2001
DEFINITION  Gossypium hirsutum cellulose synthase A4 (CesA4) gene, complete
            cds.
ACCESSION   AF413210
VERSION     AF413210.1  GI:17226293
KEYWORDS    .
SOURCE      Gossypium hirsutum (upland cotton)
  ORGANISM  Gossypium hirsutum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium.
REFERENCE   1  (bases 1 to 6886)
  AUTHORS   Kim,H.J. and Triplett,B.A.
  TITLE     Cotton fiber growth in planta and in vitro. Models for plant cell
            elongation and cell wall biogenesis
  JOURNAL   Plant Physiol. 127 (4), 1361-1366 (2001)
  MEDLINE   21608363
   PUBMED   11743074
REFERENCE   2  (bases 1 to 6886)
  AUTHORS   Kim,H.J., Williams,M.Y. and Triplett,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-AUG-2001) Fiber Bioscience, USDA-ARS, SRRC, 1100
            Robert E. Lee Blvd, New Orleans, LA 70124, USA
FEATURES             Location/Qualifiers
     source          1..6886
                     /organism="Gossypium hirsutum"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3635"
     gene            <2685..>6766
                     /gene="CesA4"
     mRNA            join(<2685..2862,3033..3106,3193..3244,3338..3482,
                     3613..3879,3992..4337,4426..4689,4778..4990,5085..5328,
                     5420..5619,5732..6085,6179..>6766)
                     /gene="CesA4"
                     /product="cellulose synthase A4"
     CDS             join(2685..2862,3033..3106,3193..3244,3338..3482,
                     3613..3879,3992..4337,4426..4689,4778..4990,5085..5328,
                     5420..5619,5732..6085,6179..6766)
                     /gene="CesA4"
                     /codon_start=1
                     /product="cellulose synthase A4"
                     /protein_id="AAL37718.1"
                     /db_xref="GI:17226294"
                     /translation="MMESGVPVCHTCGEHVGLNVNGEPFVACHECNFPICKSCFEYDL
                     KEGRKACLRCGSPYDENLLDDVEKTTGDQSTMAAHLSKSQDVGIHARHISSVSTLDSE
                     MTGDNGNPIWKNRVESWKEKKNKKKKPATTKVEREAEIPPEQQMEDKPAPDASQPLST
                     IIPIPKSRLAPYRTVIIMRLIILGLFFHYRVTHPVDSAFGLWLTSVICEIWFAFSWVL
                     DQFPKWYPVNRETYIDRLSARYDREGEPNELAAVDFFVSTVDPLKEPPLITANTVLSI
                     LALDYPVDKVSCYISDDGAAMLTFESLVETADFARKWVPFCKKFSIEPRAPEFYFSQK
                     IVYLKDKVQPSFVKERRAMKRDYEEYKIRINALVAKAQKTPEEGWTMQDGTPWPGNNP
                     RDHPGMIQVFLGYSGAHDIEGNELPRLVYVSREKRPGYQHHKQAGADNALVRVSAALT
                     NAPFILNLDCDHYVNNSKAVREAMCFLMDPQGGRDVCYVQFPQRCDGIDRSERYAKRN
                     TVFFDVNMKGRDGSQGPVYVGTGCVCNRQALYGYGPPSMPSFPKSSSSSCSCCCPGKK
                     EPKEPTELYRDAKREELDAAIFNLREIDNYDEYERSMLISQTSFEKTFGLSSVFIEST
                     LMENGGVAESANPSTLIKEAIHVIGCGYEEKTAWGKEIGWIYGSVTEDILTGFKMHCR
                     GWRSIYCMPLRPAFKGSAPINLSDRLHQVLRWALGSLEIFLSRHCPLWYGFGGGRLKW
                     LQRLAYINTSVYPFTSLPLIAYCSLPAICLLTGKFIIPTLSNLASVLYLGLFLSIIVT
                     AVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQGFLKMLAGIDTNFTVTAKAADD
                     ADFGELYIVKWTTLLIPPTTLLIVNMVGVVAGFSDALNKGYEAWGPLFGKVFFSFWVI
                     LHLYPFLKGLMGRQNRTPTIVVLWSVLLASVFSLVWVRINPFVSTADSTTVSRANSSI
                     DC"
BASE COUNT      2067 a    1237 c    1277 g    2304 t       1 others
ORIGIN

Alignment Scores:
Pred. No.:                4.41e-57        Length:          6886
Score:                    61.00           Matches:         61
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              6.24%           Indels:          0
DB:                       8               Gaps:            0

US-09-980-043-2 (1-978) x AF413210 (1-6886)

Qy        799 SerIleIleValThrAlaValLeuGluLeuArgTrpSerGlyValSerIleGluAspLeu 818
              ||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db       6221 TCCATTATCGTGACCGCTGTTCTCGAGCTCCGATGGAGTGGTGTCAGCATCGAGGACTTA 6280

Qy        819 TrpArgAsnGluGlnPheTrpValIleGlyGlyValSerAlaHisLeuPheAlaValPhe 838
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6281 TGGCGTAACGAGCAGTTTTGGGTCATCGGTGGCGTTTCAGCGCATCTCTTTGCCGTCTTC 6340
```

```
/        839 GlnGlyPheLeuLysMetLeuAlaGlyIleAspThrAsnPheThrValThrAlaLysAla 858
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
>       6341 CAAGGTTTCCTTAAGATGCTTGCAGGCATTGACACCAACTTTACTGTCACTGCCAAAGCA 6400

/        859 Ala 859
             |||
>       6401 GCT 6403


'SULT 6
'323040
)CUS        AF323040               1515 bp    mRNA    linear   PLN 19-NOV 2001
:FINITION   Zinnia elegans cellulose synthase CesA-2 mRNA, partial cds.
'CESSION    AF323040
RSION       AF323040.1  GI:13021933
YWORDS      .
)URCE       Zinnia elegans
 ORGANISM   Zinnia elegans
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; campanulids; Asterales; Asteraceae; Asteroideae;
            Heliantheae; Zinnia.
.FERENCE    1  (bases 1 to 1515)
 AUTHORS    Haigler,C.H., Ivanova-Datcheva,M., Hogan,P.S., Salnikov,V.V.,
            Hwang,S., Martin,K. and Delmer,D.P.
 TITLE      Carbon partitioning to cellulose synthesis
 JOURNAL    Plant Mol. Biol. 47 (1-2), 29-51 (2001)
 MEDLINE    21438255
  PUBMED    11554477
 FERENCE    2  (bases 1 to 1515)
 AUTHORS    Hwang,S. and Haigler,C.H.
 TITLE      Direct Submission
 JOURNAL    Submitted (22-NOV-2000) Biological Sciences, Texas Tech University,
            Box 43131, Lubbock, TX 79409, USA
 ATURES             Location/Qualifiers
    source          1. .1515
                    /organism="Zinnia elegans"
                    /mol_type="mRNA"
                    /db_xref="taxon:34245"
    CDS             <1. .1515
                    /codon_start=1
                    /product="cellulose synthase CesA-2"
                    /protein_id="AAK11589.1"
                    /db_xref="GI:13021934"
                    /translation="VREAMCFMMDPQVGRDVCYIQFPQRFDGIDRSDRYANRNTVFFD
                    VNMKGLDGIQGPVYVGTGCVFYRQALYGYGPQSLPTLPSPSSSSSCCCCGPKKPKKDL
                    EEFKRDARRDDLNAAIFNLKEIESYDDYERSLLISQMSFEKTFGMSSVFIESTLMENG
                    GLAESANPATMINEAIHVISCGYEEKTAWGKEIGWIYGSVTEDILTGFKMHCRGWRSI
                    YCMPVRPAFKGSAPINLSDRLHQVLRWALGSVEIFLSRHCPLWYGWGGGRLKLLQRLA
                    YINTIVYPFTSLPLVAYCTLPAICLLTGKFIIPTLSNIAAVWFLGLFLSIITTSVLEI
                    RWSGVSIEELWRNEQFWVIGGVSAHLFAVFQGFLKMLAGVDTNFTVTAKAADDQEFGE
                    LYMIKWTTVLIPPTTLLVLNLVGVVAGFSDALNKGYEAWGPLFGEVFFAFWVILHFTR
                    FLKSLMGRQNRTPTIVILWSVLLASVFSLVWVKIDPFVSKGDSNLTQGCIAIDC"
 SE COUNT       363 a    287 c    384 g    481 t
 IGIN

 ignment Scores:
 ed. No.:              6.03e-42        Length:         1515
 ore:                  47.00           Matches:        47
 rcent Similarity:     100.00%         Conservative:   0
 st Local Similarity:  100.00%         Mismatches:     0
 ery Match:            4.81%           Indels:         0
 :                     8               Gaps:           0

 -09-980-043-2 (1-978) x AF323040 (1-1515)

        681 GlyPheLysMetHisCysArgGlyTrpArgSerIleTyrCysMetProValArgProAla 700
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
        619 GGATTCAAGATGCATTGTAGGGGATGGAGATCAATATACTGCATGCCAGTAAGACCTGCG 678

        701 PheLysGlySerAlaProIleAsnLeuSerAspArgLeuHisGlnValLeuArgTrpAla 720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
        679 TTTAAAGGTTCGGCTCCTATCAACTTGTCAGATCGGTTGCACCAAGTTTTGAGGTGGGCC 738
```

```
Qy      721 LeuGlySerValGlu.IlePhe 727
            ||||||||||||||||||||||
Db      739 CTTGGATCGGTTGAGATTTTT 759

RESULT 7
AF323041
LOCUS       AF323041               1518 bp    mRNA    linear   PLN 19-NOV-2001
DEFINITION  Zinnia elegans cellulose synthase CesA-3 mRNA, partial cds.
ACCESSION   AF323041
VERSION     AF323041.1  GI:13021937
KEYWORDS    .
SOURCE      Zinnia elegans
  ORGANISM  Zinnia elegans
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; campanulids; Asterales; Asteraceae; Asteroideae;
            Heliantheae; Zinnia.
REFERENCE   1  (bases 1 to 1518)
  AUTHORS   Haigler,C.H., Ivanova-Datcheva,M., Hogan,P.S., Salnikov,V.V.,
            Hwang,S., Martin,K. and Delmer,D.P.
  TITLE     Carbon partitioning to cellulose synthesis
  JOURNAL   Plant Mol. Biol. 47 (1-2), 29-51 (2001)
  MEDLINE   21438255
   PUBMED   11554477
REFERENCE   2  (bases 1 to 1518)
  AUTHORS   Hwang,S. and Haigler,C.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-NOV-2000) Biological Sciences, Texas Tech University,
            Box 43131, Lubbock, TX 79409, USA
FEATURES             Location/Qualifiers
     source          1. .1518
                     /organism="Zinnia elegans"
                     /mol_type="mRNA"
                     /db_xref="taxon:34245"
     CDS             <1. .1518
                     /codon_start=1
                     /product="cellulose synthase CesA-3"
                     /protein_id="AAK11590.1"
                     /db_xref="GI:13021938"
                     /translation="VKEAMCFMMDPQVGRDVCYIQFPQRFDGIDRSDRYANRNTVFFD
                     VNMKGLDGFKGPVYVRTGCVFYRQALHGYGPQSLPTLPSPSSSSSCCCCGPKKPKKDL
                     EEFKRDARPRDDLNAAIFNLKEIESYDDYERSLLISQMSFEKTFGMSSVFIESTLMENG
                     GLARSANPATMINEAIHVISCGYEEKTAWGKEIGWIYGSVTEDILTGFKMHCRGWRSI
                     YCMSVRPAFKGSAPINLSDRLHQVLRWALGSVEIFLSRHCPLWYGWGGGRLKLLQRLA
                     YINTIVYPFTSLPLVAYCTLPAICLLTGKFIIPTLSNIAAVWFLGLFLSIITTSVLEI
                     RWSGVSIEELWRNEQFWVIGGVSAHLFAVFQGSLKMLAGVDTNFTVTAKAADDQEFGE
                     LYMIKWTTVLIPPTTLLVLNLVGVVAGFSDALNKGYEAWGPLFGKVSLRSKGDFFIC
                     RFLKGLMRQNRTPTIVILWSVLLASVFSLVWVKIDPFVSKGDSNLTQGCIAIDC"
BASE COUNT      367 a    287 c    380 g    484 t      1 others
ORIGIN

Alignment Scores:
Pred. No.:              6.04e-42        Length:         1518
Score:                  47.00           Matches:        47
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            4.81%           Indels:         0
DB:                     8               Gaps:           0

US-09-980-043-2 (1-978) x AF323041 (1-1518)

Qy      681 GlyPheLysMetHisCysArgGlyTrpArgSerIleTyrCysMetProValArgProAla 700
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      619 GGATTCAAGATGCATTGTAGGGGATGGAGATCAATATACTGCATGCCAGTAAGACCTGCG 678

Qy      701 PheLysGlySerAlaProIleAsnLeuSerAspArgLeuHisGlnValLeuArgTrpAla 720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      679 TTTAAAGGGTCGGCTCCTATCAACTTGTCAGATCGGTTGCACCAAGTTTTGAGGTGGGCC 738

Qy      721 LeuGlySerValGluIlePhe 727
            |||||||||||||||||||||
Db      739 CTTGGATCGGTTGAGATTTTT 759
```

```
RESULT 8
AF267742
LOCUS       AF267742               2958 bp    mRNA    linear   PLN 16-FEB-2001
DEFINITION  Arabidopsis thaliana cellulose synthase catalytic subunit (IRX1)
            mRNA, complete cds.
ACCESSION   AF267742
VERSION     AF267742.1  GI:12836996
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 2958)
  AUTHORS   Taylor,N.G., Laurie,S. and Turner,S.R.
  TITLE     Multiple cellulose synthase catalytic subunits are required for
            cellulose synthesis in Arabidopsis
  JOURNAL   Plant Cell 12 (12), 2529-2540 (2000)
  MEDLINE   21066939
   PUBMED   11148295
REFERENCE   2  (bases 1 to 2958)
  AUTHORS   Taylor,N.G. and Turner,S.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-MAY-2000) Biological Sciences, University of
            Manchester, 3.614 Stopford Building, Oxford Road, Manchester M13
            9PT, UK
FEATURES             Location/Qualifiers
     source          1..2958
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
     gene            1..2958
                     /gene="IRX1"
     CDS             1..2958
                     /gene="IRX1"
                     /codon_start=1
                     /product="cellulose synthase catalytic subunit"
                     /protein_id="AAK[illegible]8700.1"
                     /db_xref="GI:12836997"
                     /translation="MMESRSPICNTCGEEIGAEVNGEPFVACHECSFPICKACLEYEF
                     KEGRRICLRCGNPYDENVFDDVETKTSKTQSIVPTQTNNTSQDSGIHARHISTVSTID
                     SELNDEYGNPIWKNRVESWKDKKDKKSKKKKKDPKATKAEQDAQVPTQQHMEDMPPN
                     TESGATDVLSVVIPIPRTKITSYRIVIIMRLIILALFFNYRITHPVDSAYGLWLTSVI
                     CEIWFAVSWVLDQFPKWSPINRETYIDRLSARFEKEGEPSQLSQLAAVDFFVSTVDPLKEP
                     PLITANTVLSILALDYPVDKVSCYVSDDGAAMLSFESLVETADFARKWVPFCKKYSIE
                     PRAPEFYFSLKIDYLRDKVQPSFVKERRAMKRDYEEFKIRMNALVAKAQKTPEEGWTM
                     QDGTSWPGNNTRDHPGMIQVFLGYSGARDIEGNELPRLVYVSREKRPGYQHHKKAGAE
                     NALVRVSAVLTNAPFILNLDCDHYVNNSKAVREAMCFLMDPVVGQDVCFVQFPQRFDG
                     IDKSDRYANRNIVFFDVNMKGLDGIQGPVYVGT[illegible]VFRRQALYGYSPPSKPRILPCSS
                     SSSCCCLFKKKQPGKKFGSSTPFNLEDIDVFEMEDEVEGASSAKPVSWNYDEYDRSMLISQTSFEKT
                     FGLSTVFIESTLMENGGVPESATPASLLKEAIHVISCGYEDKTEWGKEIGWIYGSITE
                     DILTGFKMHCRGWRSIYCMPKRAAFKGSAPINLNDRLVQVLRWALGSVEIFLSRHCPL
                     WYGCSGGRLKWLQRLAYINTTVYPFTSLPLIAYCTLPAICLLTGKFIIPTLSNLASML
                     FLGLFISILTSVLVLEMKWSGITLEEWWRNEQFWVIGGVSAHLFAVFQGFLKMLAGLDT
                     NFTVTSKTADDLEFGELYIVKWTTLLIPPITLLIVNLVGVVAGFSDALNKGYEAWGPL
                     FGKVFFAFWVILHLYPFLKGLMGRQNRTPTIVIVWSILLASVFSLVWVRINPFVSKTD
                     TTSLSLNCLLIDC"
BASE COUNT      818 a    610 c    658 g    872 t
ORIGIN

Alignment Scores:
Pred. No.:              1.1e-41         Length:             2958
Score:                  47.00           Matches:               9
Percent Similarity:     100.00%         Conservative:          0
Best Local Similarity:  100.00%         Mismatches:            0
Query Match:            4.81%           Indels:                0
DB:                     8               Gaps:                  0

US-09-980-043-2 (1-978) x AF267742 (1-2958)

Qy        409 GlyAlaArgAspIleGluGlyAsnGlnProArgLeuValTyrValSerArgGluLys 428
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1252 GGTGCTCGCGACATTGAAGGAAATCAACCTCGAAGATTACGTACGTCTCTAGGGAGAAG 1311

Qy        429 ArgProGlyTyrGlnHisHisLysLysAlaGlyAlaGluAsnAlaLeuValArgValSer 448
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1312 AGACCTGGTTATCAGCATCACAAAAAGGCCGGGGCAGAGAACGCATTGGTGAGGGTGTCT 1371

Qy        449 AlaValLeuThrAsnAlaPro 455
              |||||||||||||||||||
Db       1372 GCGGTTTTAACGAATGCTCCA 1392

RESULT 9
AY099636
LOCUS       AY099636               3109 bp    mRNA    linear   PLN 06-MAY-2002
DEFINITION  Arabidopsis thaliana cellulose synthase-like protein (At4g18780)
            mRNA, complete cds.
ACCESSION   AY099636
VERSION     AY099636.1  GI:20466339
KEYWORDS    FLI_CDNA.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 3109)
  AUTHORS   Nguyen,M., Karlin-Neumann,G., Southwick,A., Lam,B., Miranda,M.,
            Palm,C.J., Bowser,L., Jones,T., Banh,J., Carninci,P., Chen,H.,
            Cheuk,R., Chung,M.K., Hayashizaki,Y., Ishida,J., Kamiya,A.,
            Kawai,J., Kim,C., Lin,J., Liu,S.X., Narusaka,M., Pham,P.K.,
            Sakano,H., Sakurai,T., Satou,M., Seki,M., Shinn,P., Yamada,K.,
            Shinozaki,K., Ecker,J., Theologis,A. and Davis,R.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-APR-2002) DNA Sequencing and Technology Center,
            Stanford University, 855 California Avenue, Palo Alto, CA 94304,
            USA
COMMENT     e-mail for correspondence: arabwsequence.stanford.edu

            RIKEN Genomic Sciences Center (GSC) members carried out the
            collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN
            Arabidopsis  Full-Length cDNA'): Seki,M., Narusaka,M., Ishida,J.,
            Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J.,
            Hayashizaki,Y. and Shinozaki,K.

            The Salk, Stanford, PGEC (SSP) Consortium members carried out the
            sequencing and annotation of the RAFL cDNAs: Nguyen,M.,
            Southwick,A., Karlin-Neumann,G., Lam,B., Miranda,M., Palm,C.J.,
            Bowser,L., Jones,T., Banh,J., Chen,H., Cheuk,R., Chung,M.K.,
            Kim,C., Lin,J., Liu,S.X., Pham,P.K., Sakano,H., Shinn,P.,
            Yamada,K., Ecker,J., Theologis,A. and Davis,R.W.

            Nguyen,M. (SSP/Stanford) and Seki,M. (RIKEN GSC) contributed
            equally to this work.  Shinozaki,K. (RIKEN GSC) and Davis,R.W.
            (SSP/Stanford) contributed equally to this work as PIs.
FEATURES             Location/Qualifiers
     source          1..3109
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
                     /chromosome="4"
                     /clone="RAFL09-65-M12 (R24904)"
                     /note="This clone is in pBluescript
                     ecotype: Columbia"
     gene            1..3109
                     /gene="At4g18780"
                     /note="synonym: F28A21.190"
     CDS             68..3025
                     /gene="At4g18780"
                     /codon_start=1
                     /product="cellulose synthase-like protein"
                     /protein_id="AAM20487.1"
                     /db_xref="GI:20466340"
                     /translation="MMESRSPICNTCGEEIGVKSNGEFFVACHECSFPICKACLEYEF
                     KEGRRICLRCGNPYDENVFDDVETKTSKTQSIVPTQTNNTSQDSGIHARHISTVSTID
                     SELNDEYGNPIWKNRVESWKDKKDKKSKKKKKDPKATKAEQHEAQIPTQQHMEDTPPN
                     TESGATDVLSVVIPIPRTKITSYRIVIIMRLIILALFFNYRITHPVDSAYGLWLTSVI
```

```
                     CEIWFAVSWVLDQFPKWSPINRETYIDRLSARFEREGEQSQLAAVDFFVSTVDPLKEP
                     PLITANTVLSILALDYPVDKVSCYVSDDGAAMLSFESLVETADFARKWVPFCKKYSIE
                     PRAPEFYFSLKIDYLRDKVQPSFVKERRAMKRDYEEFKIRMNALVAKAQKTPEEGWTM
                     QDGTSWPGNNTRDHPGMIQVFLGYSGARDIEGNELPRLVYVSREKRPGYQHHKKAGAE
                     NALVRVSAVLTNAPFILNLDCDHYVNNSKAVREAMCFLMDPVVGQDVCFVQFPQRFDG
                     IDKSDRYANRNIVFFDVNMRGLDGIQGPVYVGTGTVFRRQALYGYSPPSKPRILPQSS
                     SSSCCCLTKKKQPQDPSEIYKDAKREELDAAIFNLGDLDNYDEYDRSMLISQTSFEKT
                     FGLSTVFIESTLMENGGVPDSVNPSTLIKEAIHVISCGYEEKTEWGKEIGWIYGSITE
                     DILTGFKMHCRGWRSIYCMPLRPAFKGSAPINLSDRLHQVLRWALGSVEIFLSRHCPL
                     WYGCSGGRLKLLQRLAYINTIVYPFTSLPLVAYCTLPAICLLTGKFIIPTLSNLASML
                     FLGLFISIILTSVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQGFLKMLAGLDT
                     NFTVTSKTADDLEFGELYIVKWTTLLIPPTSLLIINLVGVVAGFSDALNKGYEAWGPL
                     FGKVFFAFWVILHLYPFLKGLMGRQNRTPTIVILWSILLASVFSLVWVRINPFVSKTD
                     TTSLSLNCLLIDC"
ASE COUNT      859 a    631 c    687 g    932 t
RIGIN

ignment Scores:
ed. No.:              1.15e-41       Length:        3109
core:                 47.00          Matches:       47
ercent Similarity:    100.00%        Conservative:  0
est Local Similarity: 100.00%        Mismatches:    0
uery Match:           4.81%          Indels:        0
B:                    8              Gaps:          0

3-09-980-043-2 (1-978) x AY099636 (1-3109)

,        409 GlyAlaArgAspIleGluGlyAsnGluLeuProArgLeuValTyrValSerArgGluLys 428
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
>       1319 GGTGCTCGCGACATTGAAGGAAATGAACTTCCAAGATTAGTTTACGTCTCTAGAGAGAAG 1378

         429 ArgProGlyTyrGlnHisHisLysLysAlaGlyAlaGluAsnAlaLeuValArgValSer 448
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
        1379 AGACCTGGTTATCAGCATCACAAAAAGGCCGGGGCAGAGAACGCATTGGTGAGGGTGTCT 1438

         449 AlaValLeuThrAsnAlaPro 455
             |||||||||||||||||||
        1439 GCGGTTTTAACGAATGCTCCA 1459

SULT 10
197911
CUS       AF197911                 1010 bp    DNA     linear   PLN 02-AUG-2000
FINITION  Populus tremuloides cellulose synthase (CesA) gene, promoter and
          partial cds.
CESSION   AF197911
RSION     AF197911.1  GI:9651405
YWORDS    .
URCE      Populus tremuloides (quaking aspen)
ORGANISM  Populus tremuloides
          Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
          Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
          rosids; eurosids I; Malpighiales; Salicaceae; Populus.
FERENCE   1  (bases 1 to 1010)
AUTHORS   Wu,L., Joshi,C.P. and Chiang,V.L.
TITLE     A xylem-specific cellulose synthase gene from aspen (Populus
          tremuloides) is responsive to mechanical stress
JOURNAL   Plant J. 22 (6), 495-502 (2000)
MEDLINE   20344632
 PUBMED   10886769
FERENCE   2  (bases 1 to 1010)
AUTHORS   Wu,L., Joshi,C.P. and Chiang,V.L.
TITLE     Direct Submission
JOURNAL   Submitted (22-OCT-1999) Forestry, Michigan Tech University, 1400,
          Townsend Drive, Houghton, MI 49931, USA
ATURES              Location/Qualifiers
   source           1. .1010
                    /organism="Populus tremuloides"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:3693"
                    /tissue_type="xylem"
   gene             <1. .>1010
                    /gene="CesA"
   misc_feature     1. .840
                    /gene="CesA"
                    /note="contains promoter and 5'UTR"
   mRNA             <841. .>1010
                    /gene="CesA"
                    /product="cellulose synthase"
   CDS              841. .>1010
                    /gene="CesA"
                    /codon_start=1
                    /product="cellulose synthase"
                    /protein_id="AAF91155.1"
                    /db_xref="GI:9651406"
                    /translation="MMESGAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEI
                    KEGRKVCPQCS"
BASE COUNT      289 a    21 c    216 g    293 t
ORIGIN

Alignment Scores:
Pred. No.:                1.15e-41       Length:        1010
Score:                    44.00          Matches:       44
Percent Similarity:       100.00%        Conservative:  0
Best Local Similarity:    100.00%        Mismatches:    0
Query Match:              4.63%          Indels:        0
DB:                       8              Gaps:          0

US 09-980-043-2 (1-978) x AF197911 (1-1010)

Qy          1 MetMetGluSerGlyAlaProIleCysHisThrCysGlyGluGlnValGlyHisAspAla 20
              |||||||| ||| ||| ||| |||| |||| |||||||||||||||||||||| || ||
Db        841 ATGATGGAATCTGGTGCTCCGATATGCCATACTCTTGTGGTGAACAGGTGGGGCATGATGCA 900

Qy         21 AsnGlyGluLeuProValAlaCysHisGluCysSerTyrProMetCysLysSerCysPhe 40
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AATGGGGAGCTATTTGTGGCTTGTCATGAATGTACCTATCCCATGTGCAAGTCTTGTTTC 960

Qy         41 GluPheGluIle 43
              |||||||||||
Db        961 GAGTTTGAAATC 973

RESULT 11
ATF28A21/c
LOCUS       ATF28A21               94301 bp    DNA     linear   PLN 08-SEP 1999
DEFINITION  Arabidopsis thaliana DNA chromosome 4, BAC clone F28A21 (ESSA
            project).
ACCESSION   AL035526
VERSION     AL035526.1  GI:4[illegible]
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Bevan,M., Mueller,M.W., Murdlein,A., Felber,R., Bancroft,I.,
            Mewes,H.W., Mayer,K.F.X., Lemcke,K. and Schueller,C.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 94301)
  AUTHORS   EU Arabidopsis sequencing project
  TITLE     Direct Submission
  JOURNAL   Submitted (08-SEP-1999) MIPS, at the Max Planck-Institut fuer
            Biochemie, Am Klopferspitz 18a, D 82152 Martinsried, FRG, E-mail:
            schuelle@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de Project
            Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
            Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
            E-mail: michael.bevan@bbsrc.ac.uk
COMMENT     Information on performance of analysis and a more detailed
            annotation of this entry and other sequences of chromosomes 3, 4
            and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/.
FEATURES            Location/Qualifiers
     source         1. .94301
                    /organism="Arabidopsis thaliana"
                    /mol_type="genomic DNA"
                    /variety="Columbia"
                    /db_xref="taxon:3702"
```

```
                /chromosome="4"
gene            9652. .10227
                /gene="F28A21.20"
CDS             9652. .10227
                /gene="F28A21.20"
                /note="similarity to predicted proteins, Arabidopsis
                thaliana
                contains EST gb:T46643, R29928"
                /codon_start=1
                /product="putative protein"
                /protein_id="CAB37446.1"
                /db_xref="GI:4539380"
                /translation="MSSDRHTPTKDPPDHPSSS MHHKQPLPPQPQQCPLSRYESQKRR
                DWNTFVQYLKSQNPPLMMSCFDYTHVLSFLRYL CPGKTKVHHQACVPPGQPDPPGPC
                TCPLKQAWGSLDALIGFLRAAYEHGGGSPDTN FANGSIRVHLREVRESQAKARGIP
                YRKKKRRKTKNEVVVVRKDVANSSTPNQSFT"
exon            9652. .10227
                /gene="F28A21.20"
                /number=1
gene            13217 .13711
                /gene="F28A21.30"
CDS             13217. .13711
                /gene="F28A21.30"
                /note="similarity to various predicted proteins,
                Arabidopsis thaliana"
                /codon_start=1
                /product="putative protein"
                /protein_id="CAB37447.1"
                /db_xref="GI:4539381"
                /translation="MESSKQKRCRSSVVETIEAI LPLVWSILRSFDKPQAYQRFVKSC
                TMRSGGGGGKGGEGKGSVRDVTLVSGFPADFSTE RLEELDDESHVMVVSIIGGNHRLV
                NYKSKTKVVVASPEDMAKKTVVVESYVVDVPEGT EEDTIFFVDNIIRYNLTSLAKLTK
                KMMK"
exon            13217. .13711
                /gene="F28A21.30"
                /number=1
gene            16036. .17454
                /gene="F28A21.40"
CDS             16036. .17454
                /gene="F28A21.40"
                /note="similarity to predicted protein, Arabidopsis
                thaliana"
                /codon_start=1
                /product="putative protein"
                /protein_id="CAB37448.1"
                /db_xref="GI:4539382"
                /translation="MEGKKLNLYAPLPSIRRIP MIERSSELENKKTTITRPELRSCP
                STKQETPVFVLPDQSFDHLTEPASIPFMWEQIF GKPKDDMATLIQESGLLETDDEEED
                EVDEDLDTVSSNVSFSVN ST GVSFTEKTGER DEKSRESLDLMMSRFLPAAKAM
                ALQTHQKHQSSYNSSF LNEALVAF ANAEHEHFAIVQSLYDDLNIDDDT
                EDDENDEDGDDYGHVD FAKDSEKFLNPVPLDSVSNRSMK
                HSGDQTSLPNWSTRRL EKPESFKRLNRGISKSQELYPT
                RTRREVILPNYSNSGE SRPLVEEVNRRRNNKSGNLQQT
                SSEYTPALSPPPLPST TIVPVLGQVGIKKLNQEVLESTKWE
                TIVKTSYVHNDHARYSQVLITL"
exon            16036. .17454
                /gene="F28A21.40"
                /number=1
gene            19501. .22435
                /gene="F28A21.50"
CDS             join(<19501. .19652,19742. .19813,19902. .20045,20144. .20287,
                20375. .20469,20559. .20987,21071. .21261,21342. .21474,
                21531. .21709,21893. .22084,22178. .22435)
                /gene="F28A21.50"
                /codon_start=1
                /product="putative protein (fragment)"
                /protein_id="CAB37449.1"
                /db_xref="GI:4539383"
                /translation="FVLNSGSALLKFRARVNSDHGTLANWNVSGINDLCYWSGVTCV
                DGKVQILDLSGYSLEGTLAPELSQLSDLRSLILSRNHFSGGIPKEYGSFENLEVLDLR
                ENDLSGQIPPELSNGLSLKHLDLSGNKFSDDMR KIVRLQSSYEVRLKKSPKLSPLAV
                LGCINRKLGHCVSRNRI QVKKVEAIVFRIKATSPRFLKAFPSFLEETDIYKRRELLE
                ETSNLAAEPAPSAPSPSPGI ITEASPRSSGSPPAVTNAKKRRPPLVPPVPSPDKGSTS
                PDISKNQPQDNKQSKGSKHVWLYVVIAVASFVGLLIIVAVIFFCRKRAVKSIGPWKTG
                LSGQLQKAFVTGVPKLNRSELETACEDFSNIIETFDGYTVYKGTLSSGVEIAVASTAI
                AESKEWTRAMEMAYRRKIDTLSRINHKNFVNLIGYCEEDDPFNRMMVFEYAPNGTLFE
                HLHVNNVFRLSDKETEHLDWSARMRIIMGTAYCLQHMHGMNPPMAHTDFNSSEIYLTD
                DYAAKVSEIPFNLEARLNPKKHVSGDLEQTSLLLPPEPEANVHSFGVLMLEIISGKLS
                FSDEYGSIEQWASKYLEKDDLGEMIDPSLKTFKEEELEVICDVIRECLKTEQRQRPSM
                KDVAEQLKQVINITPEKATPRSSPLWWAELEILSSEAT"
exon            19501. .19652
                /gene="F28A21.50"
                /number=1
intron          19653. .19741
                /gene="F28A21.50"
                /number=1
exon            19742. .19813
                /gene="F28A21.50"
                /number=2
intron          19814. .19901
                /gene="F28A21.50"
                /number=2
exon            19902. .20045
                /gene="F28A21.50"
                /number=3
intron          20046. .20143
                /gene="F28A21.50"
                /number=3
exon            20144. .20287
                /gene="F28A21.50"
                /number=4
intron          20288. .20374
                /gene="F28A21.50"
                /number=4
exon            20375. .20469
                /gene="F28A21.50"
                /number=5
intron          20470. .20558
                /gene="F28A21.50"
                /number=5
exon            20559. .20987
                /gene="F28A21.50"
                /number=6
intron          20988. .21070
                /gene="F28A21.50"
                /number=6
exon            21071. .21261
                /gene="F28A21.50"
                /number=7
intron          21262. .21341
                /gene="F28A21.50"
                /number=7
exon            21342. .21474
                /gene="F28A21.50"
                /number=8
intron          21475. .21530
                /gene="F28A21.50"
                /number=8
exon            21531. .21709
                /gene="F28A21.50"
                /number=9
intron          21710. .21892
                /gene="F28A21.50"
                /number=9
exon            21893. .22084
                /gene="F28A21.50"
                /number=10
intron          22085. .22177
                /gene="F28A21.50"
                /number=10
exon            22178. .22435
                /gene="F28A21.50"
                /number=11
gene            22901. .24271
                /gene="F28A21.60"
gene            complement(join(22901. .23176,23858. .24271))
```

```
                /gene="F28A21.60"
   CDS          complement(join(22901. .23176,23858. .24271))
                /gene="F28A21.60"
                /note="similarity to transcription factor OBF3.1, Zea
                mays, PIR2:S33223"
                /codon_start=1
                /product="putative protein"
                /protein_id="CAB37450.1"
                /db_xref="GI:4539384"
                /translation="MRNLVEEKFLEFYESWVIQLELYLHQLLIAHNNNTMSETELRHL
                ISKLTTHHKAYYTAKWAAIREDVLAFFGSVWLNPLENACSWLTGWKPSMVFRMVDRLR
                KSRVVLVEAQVKKLEELRVKTKFDEQKIEREMERYQVAMADRKMVELARLGCHVGGES
                VMVVEAAVRGLSMGLEKMVKAADCVRLKTLKGILDILTPPQCVEFLAAAATFQVQLRR
                WGNRRHYVTHS"
   exon         complement(22901. .23176)
                /gene="F28A21.60"
                /number=1
   intron       complement(23177. .23857)
                /gene="F28A21.60"
                /number=1
   exon         complement(23858. .24271)

ignment Scores:
ed. No.:            7.46e-36        Length:         94301
ore:                43.00           Matches:        43
rcent Similarity:   100.00%         Conservative:   0
st Local Similarity: 100.00%        Mismatches:     0
ery Match:          4.40%           Indels:         0
                    8               Gaps:           0

-09-980-043-2 (1-978) x ATF28A21 (1-94301)

      805 ValLeuGluLeuArgTrpSerGlyValSerIleGluAspLeuTrpArgAsnGluGlnPhe 824
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
    72231 GTCCTCGAGCTTCGATGGAGCGGAGTCAGTATCGAAGACTTATGGAGAAACGAACAGTTT 72172

      825 TrpValIleGlyGlyValSerAlaHisLeuPheAlaValPheGlnGlyPheLeuLysMet 844
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
    72171 TGGGTTATTGGAGGTGTCTCAGCTCATCTCTTTGCCGTTTTCCAAGGATTCCTCAAAATG 72112

      845 LeuAlaGly 847
          |||||||||
    72111 CTCGCTGGT 72103

SULT 12
CHRIV49/c
CUS        ATCHRIV49             199075 bp    DNA     linear   PLN 16-MAR-2000
FINITION   Arabidopsis thaliana DNA chromosome 4, contig fragment No. 49.
CESSION    AL161549
RSION      AL161549.2  GI:7268649
YWORDS     .
JRCE       Arabidopsis thaliana (thale cress)
ORGANISM   Arabidopsis thaliana
           Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
           Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
           rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
FERENCE    1  (bases 1 to 28924)
AUTHORS    Hilbert,H., Braun,M., Holzer,E., Brandt,A., Duesterhoeft,A.,
           Mewes,H.W., Lemcke,K. and Mayer,K.F.X.
JOURNAL    Unpublished
FERENCE    2  (bases 19282 to 113583)
AUTHORS    Mueller,M.W., Muendlein,A., Felber,R., Mewes,H.W., Lemcke,K. and
           Mayer,K.F.X.
JOURNAL    Unpublished
FERENCE    3  (bases 113371 to 118148)
AUTHORS    Robben,J., Grymonprez,B., Volckaert,G., Mewes,H.W., Lemcke,K. and
           Mayer,K.F.X.
JOURNAL    Unpublished
FERENCE    4  (bases 117865 to 199075)
AUTHORS    Pohl,T., Weizenegger,T., Mewes,H.W., Lemcke,K. and Mayer,K.F.X.
JOURNAL    Unpublished
FERENCE    5  (bases 1 to 199075)
AUTHORS    EU Arabidopsis sequencing,project.

  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer
            Biochemie, Am Klopferspitz 18a, D 82152 Martinsried, FRG, E-mail:
            lemcke@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de Project
            Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
            Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
            E-mail: michael.bevan@bbsrc.ac.uk
COMMENT     Information on performance of analysis and a more detailed
            annotation of this entry and other sequences of chromosome 1, 4
            and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/
            this fragment has an overlap with ATCHRIV48 at the 5' end and an
            overlap with ATCHRIV50 at the 3' end.
FEATURES              Location/Qualifiers
     source           1. .[illegible]
                      /organism="Arabidopsis thaliana"
                      /mol_type="genomic DNA"
                      /strain="Columbia"
                      /db_xref="taxon:3702"
                      /chromosome="4"
     gene             9577. .12678
                      /gene="AT4g18560"
     CDS              join(9577. .9906,9988. .10809,10980. .11111,11160. .11377,
                      11463. .11574,12313. .12447,12541. .12672)
                      /gene="AT4g18560"
                      /note="similarity to pherophorin-S, Volvox carteri,
                      PIR:T10708"
                      /codon_start=1
                      /product="pherophorin like protein"
                      /protein_id="CAB78858.1"
                      /db_xref="GI:7268650"
                      /translation="MVAGKVRVTMPHKSPSTKKTKDMPSPLPLPPPPPPLKPPSSG
                      SATTKPPKPSKPGPTRSPGVYFPPASAQVHATAAAASHNGVVSFLRRQVEELRREEA
                      LLKTENLHADKNGEIDELFKETARLAEDNERLRREFDRSEEMREPETREKEMEAFI
                      VELRKLVSSESDDHALSVSQRFQGLMDVSAKSNLIRSLKRVGSLRNLPEPITNQENTN
                      KSISSSGIADGDIYRKDEIESYSRGSNSEELTESSSLSTVPSRVPRVPKPPPKRSISL
                      GDSTENRADPPPQKSIPPPPPPPPPPLLQQPPPPPSVSKAPPPPPPPPPPKSLSIASA
                      KVRRVPEVVEFYHSLMRRDSTNSRRDSTGGGNAAAEAILANSNARDMIGEIENRSVYL
                      LAIKTDVETQGDFIRFLIKEVGNAAFSDIEDVVPFVKWLDDELSYLFKVCKFVVVSLK
                      VDERAVLKHFEWPEQKADALRRAAFCYFDLKKLISEASRFREDFRQSSSSALKKMQAL
                      FEKLEHGVYSLSRMRESAATKFKSPQIPVDWMLETGITSQIKLASVKLAMKYMKRVSA
                      ELEAIEGAPEEEELIVQAVRFAFRVHQFAGGFDAETMKAFEELRDKARSCHVQCQSQ
                      THQHKLCFRSTPC"
     exon             9577. .9906
                      /gene="AT4g18560"
                      /number=1
     gene             9612. .11099
                      /gene="AT4g18570"
     gene             complement(join(9612. .9715,10483. .10747,10973. .11099))
                      /gene="AT4g18570"
     CDS              complement(join(9612. .9715,10483. .10747,10973. .11099))
                      /gene="AT4g18570"
                      /codon_start=1
                      /product="hypothetical protein"
                      /protein_id="CAB78859.1"
                      /db_xref="GI:7268651"
                      /translation="MSFNAAFPTSLIKNLMKSPCVSTSVFICNQVLLPRRRFHRRWNL
                      FSSLWSLIASTSDKTQPLPVLPSLAPTQCLNSSAAVVEEAEEELWKQTEVVVVVVEAV
                      AEVEEEEAEEISAVAVAEPEEPDGLAGGAGGGGRGKGEGMSLVPLVEGDL"
     exon             complement(9612. .9715)
                      /gene="AT4g18570"
                      /number=1
     intron           complement(9716. .10482)
                      /gene="AT4g18570"
                      /number=1
     intron           9907. .9957
                      /gene="AT4g18570"
                      /number=1
     exon             9958. .10809
                      /gene="AT4g18570"
                      /number=2
     exon             complement(10483. .10747)
                      /gene="AT4g18570"
```

```
                /number=2
intron          complement(10748. .10972)
                /gene="AT4g18570"
                /number=2
intron          10810. .10979
                /gene="AT4g18570"
                /number=2
exon            complement(10973. .11069)
                /gene="AT4g18570"
                /number=3
exon            10980. .11111
                /gene="AT4g18570"
                /number=3
intron          11112 .11159
                /gene="AT4g18560"
                /number=3
exon            11160. .11377
                /gene="AT4g18560"
                /number=4
intron          11378. .11462
                /gene="AT4g18560"
                /number=4
exon            11463. .11574
                /gene="AT4g18560"
                /number=5
intron          11575. .12309
                /gene="AT4g18560"
                /number=5
exon            12310. .12447
                /gene="AT4g18560"
                /number=6
intron          12448. .12540
                /gene="AT4g18560"
                /number=6
exon            12541. .12672
                /gene="AT4g18560"
                /number=7
gene            12991. .13662
                /gene="AT4g18580"
gene            complement(join(13077,13264..13382. .13507,
                13596. .13662))
                /gene="AT4g18580"
CDS             complement(join(13077,13264..13382. .13507,
                13596. .13662))
                /gene="AT4g18580"
                /note="contains EST gb:AI998801.1, AI148148"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="AAF
                /db_xref="GI:
                /translation="MA
                PVDTFDFYNGSFDEKG
                GDDESKAKDAGYSSDE
exon            complement(12991. .13
                /gene="AT4g18580"
                /number=1
intron          complement(13078. .13
                /gene="AT4g18580"
                /number=1
exon            complement(13176. .13
                /gene="AT4g18580"
                /number=2
intron          complement(13265. .13
                /gene="AT4g18580"
                /number=2
exon            complement(13382. .13
                /gene="AT4g18580"
                /number=3
intron          complement(13508. .13
                /gene="AT4g18580"
                /number=3
exon            complement(13596. .13
                /gene="AT4g18580"
                /number=4
gene            14662. .20681
                /gene="AT4g18590"
CDS             join(14662. .14917,15014. .15052,16278. .16535,16652. .16752,
                17087. .17423,17520. .17854,18312. .18366,19447. .19527,
                19615. .20681)
                /gene="AT4g18590"
                /note="overlap to next genmodel
                similarity to 18.3K protein precursor, pollen, Zea mays,
                PIR2:JQ1107
                Contains Pollen proteins Ole e I family signature
                AA309-318
                contains EST gb:AI999674.1, AA721906"
                /codon_start=1
                /product="pollen-specific protein-like"
                /protein_id="CAB78861.1"
                /db_xref="GI:7268653"
                /translation="MDTSSPAAFVNGALLRRYIGQKVRAVIQVIRSDVGSVIGKSTDD
                QQIVVKGSPQPPLTTYLEVIGIAETDNTIRAEVWTNFGDSFDVQNYNELCKLANDRRK
                AMTDLQMEVEVDTNSSLQESLPKPQVMYRCKKCRRIVAIEENIVPHEPGKGEECFAWK
                KRSGNSEQVQCSSIFVEPMKWMQTIHDGMVEEKLLCFGCNGRLGYFNWAGMQCSCGAW
                LKITDGFNFMILYKTELCGSCELVVLFSNISNKENDYKREEVSKRRPESFNILKQMAS
                KAIFFFFVSAVCLSSLAGVAIADADDFDRFQIQGSVYCDTCRVQFVTRLSKFLEGAKV
                KLECRSRTNGTVTLTKEAVTDKTGSYRMEVTGDHEEEVCELVLVESPDSGCSDVSKEA
                YLRNAAKISLTANDGIVSHETRIVNPLGFMVQTPSAECPAAFKELGIVPDGLLFGWFG
                GQARMISETDGRSVTTHGAATCISGSFKEGYVESGTSAKARYAAAGSFGRIFSLGTGS
                SGKHLLCEPVSVVKLGTPEPVIFSSKTASVEDRSVVIVSLGKVLLCGLVSVVELEASE

Alignment Scores:
Pred. No.:              1.45e-35        Length:         199075
Score:                  43.00           Matches:        43
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            4.40%           Indels:         0
DB:                     8               Gaps:           0

US-09-980-043-2 (1-978) x ATCHRIV49 (1-199075)

Qy        805 ValLeuGluLeuArgTrpSerGlyValSerIleGluAspLeuTrpArgAsnGluGlnPhe 824
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      91512 GTCCTCGAGCTTCGATGGAGCGGAGTCAGTATCGAAGACTTATGGAGAAACGAACAGTTT 91453

Qy        825 TrpValIleGlyGlyValSerAlaHisLeuPheAlaValPheGlnGlyPheLeuLysMet 844
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      91452 TGGGTTATTGGAGGTGTCTCAGCTCATCTCTTTGCCGTTTTCCAAGGATTCCTCAAAATG 91393

Qy        845 LeuAlaGly 847
              |||||||||.
Db      91392 CTCGCTGGT 91384

RESULT 15
AR267561
LOCUS       AR267561               1741 bp    DNA     linear   PAT 10-APR-2003
DEFINITION  Sequence 13 from patent US 6495740.
ACCESSION   AR267561
VERSION     AR267561.1  GI:29697637
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1741)
  AUTHORS   Arioli,A., Williamson,R.E., Betzner,A.S. and Peng,L.
  TITLE     Manipulation of cellulose and/or .beta.-1,4-Glucan
  JOURNAL   Patent: US 6495740-A 13 17-DEC-2002;
FEATURES             Location/Qualifiers
     source          1. .1741
                     /organism="unknown"
BASE COUNT      440 a    393 c    486 g    416 t
ORIGIN

Alignment Scores:
Pred. No.:              4.9fe-34        Length:         1741
Score:                  40.00           Matches:        40
```

```
rcent Similarity:     100.00%        Conservative:    0
t Local Similarity:   100.00%        Mismatches:      0
ry Match:             4.09%          Indels:          0
                      6              Gaps:            0

 09-980-043-2 (1-978) x AR267561 (1-1741)

       256 PheValSerThrValAspProLeuLysGluProProLeuIleThrAlaAsnThrValLeu 275
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
      1160 TTTGTCAGTACGGTGGATCCACTAAAGGAACCTCCTCTGATCACAGCAAACACTGTTTTG 1219

       276 SerIleLeuAlaValAspTyrProValAspLysValSerCysTyrValSerAspAspGly 295
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
      1220 TCCATTCTGGCTGTGGATTACCCTGTTGACAAAGTGTCATGCTATGTTTCTGACGATGGT 1279

 ULT 14
 30950
 CUS       AX030950              1741 bp    DNA     linear   PAT 24-NOV-2000
 INITION   Sequence 13 from Patent WO9800549.
 CESSION   AX030950 AX030964
  SION     AX030950.1  GI:10278353
 WORDS     .
 URCE      Oryza sativa
 ORGANISM  Oryza sativa
           Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
           Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
           Ehrhartoideae; Oryzeae; Oryza.
 ERENCE    1
 AUTHORS   Williamson,R.E., Peng,L., Arioli,A. and Betzner,A.S.
 ITLE      Manipulation of cellulose and/or beta -1,4-glucan
 OURNAL    Patent: WO 9800549-A 13 08-JAN-1998;
           WILLIAMSON RICHARD EDWARD (AU) ; PENG LIANGCAI (AU) ; ARIOLI
           ANTONIO (AU) ; UNIV AUSTRALIAN (AU) ; BETZNER ANDREAS STEFAN (AU) ;
           COMMW SCIENT IND RES ORG (AU)
 MENT      On Oct 15, 2002 this sequence version replaced gi:10278367.
 ATURES              Location/Qualifiers
    source           1. .1741
                     /organism="Oryza sativa"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4530"
                     /clone="S0542"
    CDS              101. .>1741
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAC09629.1"
                     /db_xref="GI:10278354"
                     /db_xref="REMTREMBL:CAC09629"
                     /translation="MAANAGMVAGSRNRNEFVMIRPDGDAPPPAKPGKSVNGQVCQIC
                     GDTVGVSATGDVFVACNECAFPVCRPCYEYERKEGNQCCPQCKTRYKRHKGCPRVQGD
                     EEEEDVDDLDNEFHYKHGNGKGPEWQIQRQGEDVDLSSSSRHEQHRIPRLTSGQQISG
                     EIPDASPDRHSIRSGTSSYVDPSVPVPVRIVDPSKDLNSYGINSVDWQERVASWRNKQ
                     DKNMMQVANKYPEARGGDMEGTGSNGEDIQMVDDARLPLSRIVPIPSNQLNLYRIVII
                     LRLIILMFFFQYRVTHPVRDAYGLWLVSVICEIWLPLSWLLDQFPKWYPINRETYLDR
                     LALRYDREGEPSQLAPIDVFVSTVDPLKEPPLITANTVLSILAVDYPVDKVSCYVSDD
                     GSAMLTFEALSETAEFARKWVPFCKKHNIEPRAPEFYFAQKIDYLKDKIQPSFVKERR
                     AMKREYEEFKVRINALVAKAQKVPEEGWTMADGTAWPGNNPRDHPGMIQVFLGHSGGL
                     DTDGNELPRLVYVSREKRPGFQHHKKAGAMNALIRVSAV"
 E COUNT       440 a    399 c    486 g    416 t
 GIN

 gnment Scores:
 d. No.:               4.96e-34       Length:          1741
 ore:                  40.00          Matches:         40
 cent Similarity:      100.00%        Conservative:    0
 t Local Similarity:   100.00%        Mismatches:      0
 ry Match:             4.09%          Indels:          0
                       6              Gaps:            0

 09-980-043-2 (1-978) x AX030950 (1-1741)

       256 PheValSerThrValAspProLeuLysGluProProLeuIleThrAlaAsnThrValLeu 275
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
      1160 TTTGTCAGTACGGTGGATCCACTAAAGGAACCTCCTCTGATCACAGCAAACACTGTTTTG 1219
```

```
Qy        276 SerIleLeuAlaVa AspTyrProValAspLysValSerCysTyrValSerAspAspGly 295
              |||:|||: ||| .|| ||| ||| :|||.|||||:||||:| |||:|||
Db       1220 TCCATTCTGGCTGT GATTACCCTGTTGACAAAGTGTCATGCTATGTTTCTGACGATGGT 1279

RESULT 15
BD022680
LOCUS       BD022680               1741 bp    DNA     linear   PAT 27-AUG 2002
DEFINITION  Manipulation of cellulose and/or beta-1,4-glucan.
ACCESSION   BD022680
VERSION     BD022680.1  GI:[illegible]
KEYWORDS    JP 2001510366-A/1.
SOURCE      Oryza sativa
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 1741)
  AUTHORS   Arioli,A., W[illegible] F., [illegible],A.S. and Peng,L.
  TITLE     Manipulation of cellulose and/or beta 1,4-glucan
  JOURNAL   Patent: JP 2001510366-A/1 31-JUL-2001;
            THE AUSTRALIAN NATIONAL UNIVERSITY, COMMONWEALTH SCIENTIFIC AND
            INDUSTRIAL RESEARCH ORGANIZATION
COMMENT     OS   Oryza sativa (rice)
            PN   JP 2001510366-A/1
            PD   31-JUL-2001
            PF   24-JUN-1997 JP 1998503665
            PR   27-JUN-1996 AU PO0894;
            PI   ANTONIO ARIOLI,RICHARD EDWARD WILLIAMSON,ANDREAS STEFAN BE
            PI   LYONKAI PENN
            PC   C12N15/09,A[illegible],C07K14/40,C12N1/15,C12N1/19,C12N1/21,C12N5/
            PC   10,C12N5/1[illegible]
            PC   C12N9/10,C12N15/00,C12P[illegible],C12N1/00
            CC   Strandedness: Single;
            CC   Topology: Linear;
            FH   Key             Location/Qualifiers
            FT   CDS             101. .1741
FEATURES             Location/Qualifiers
     source          1. .1741
                     /organism="Oryza sativa"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4530"
BASE COUNT      440 a    399 c    486 g    416 t
ORIGIN

Alignment Scores:
Pred. No.:              4.96e-34       Length:          1741
Score:                  40.00          Matches:         40
Percent Similarity:     100.00%        Conservative:    0
Best Local Similarity:  100.00%        Mismatches:      0
Query Match:            4.09%          Indels:          0
DB:                     6              Gaps:            0

US-09-980-043-2 (1-978) x BD022680 (1-1741)

Qy        256 PheValSerThrValAspProLeuLysGluProProLeuIleThrAlaAsnThrValLeu 275
              |.|||.||||.||.||  ||| ||| .||| ||||||||||||||||| ||||:|||
Db       1160 TTTGTCAGTACGGT GATCCACTAAAGGAACCTCCTCTGATCACAGCAAACACTGTTTTG 1219

Qy        276 SerIleLeuAlaVa AspTyrProValAspLysValSerCysTyrValSerAspAspGly 295
              |||. |||: ||| || ||| :|||:| ||| .||||:||||| |||:| |||  |
Db       1220 TCCATTCTGGCTGT GATTACCCTGTTGACAAAGTGTCATGCTATGTTTCTGACGATGGT 1279


Search completed: November  1, 2003, 03:04:20
Job time : 10727.2 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 30, 2003, 23:15:30 ; Search time 170.895 Seconds
(without alignments)
8347.518 Million cell updates/sec

Title: US-09-980-043-1
Perfect score: 3232
Sequence: 1 gtgaacccacgcgtgtcgtct...aaaaaaaaaaaaa 3232

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 5e9978 seqs, 220691568 residues

Word size 6

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	30	0.9	81	3	US-09-014-416-57	Sequence 57, Appl
	2	30	0.9	127	3	US-09-014-416-59	Sequence 59, Appl
	3	30	0.9	176	3	US-09-014-416-63	Sequence 63, Appl
c	4	30	0.9	184	3	US-09-014-416-60	Sequence 60, Appl
c	5	30	0.9	200	3	US-09-014-416-64	Sequence 64, Appl
c	6	30	0.9	9598	4	US-09-014-416-4	Sequence 4, Appli
c	7	30	0.9	9599	3	US-09-014-416-2	Sequence 2, Appli
c	8	30	0.9	9599	3	US-09-014-416-6	Sequence 6, Appli
	9	29	0.9	44	1	US-08-664-596B-9	Sequence 9, Appli
c	10	29	0.9	356	2	US-08-520-678A-22	Sequence 22, Appl
c	11	29	0.9	356	3	US-08-897-126-22	Sequence 22, Appl
c	12	29	0.9	537	4	US-09-720-201A-4	Sequence 4, Appli
	13	29	0.9	572	3	US-09-357-241-7	Sequence 7, Appli
c	14	29	0.9	599	3	US-09-329-111-147	Sequence 147, App
c	15	29	0.9	603	3	US-09-385-982-293	Sequence 293, App
c	16	29	0.9	612	3	US-09-385-982-280	Sequence 280, App
c	17	29	0.9	644	4	US-09-720-201A-6	Sequence 6, Appli
	18	29	0.9	790	3	US-09-363-970-4	Sequence 4, Appli
	19	29	0.9	1192	4	US-09-439-954-23	Sequence 23, Appl
	20	29	0.9	1212	4	US-09-182-145-34	Sequence 34, Appl
c	21	29	0.9	1212	4	US-09-182-145-35	Sequence 35, Appl
	22	29	0.9	1558	1	US-08-467-607-2	Sequence 2, Appli
	23	29	0.9	1558	2	US-08-469-362-2	Sequence 2, Appli
	24	29	0.9	1558	2	US-08-850-392-2	Sequence 2, Appli
	25	29	0.9	2280	3	US-08-813-150-1	Sequence 1, Appli
	26	29	0.9	2280	4	US-09-546-553-1	Sequence 1, Appli
	27	29	0.9	2351	1	US-08-168-091A-1	Sequence 1, Appli
	28	29	0.9	3328	3	US-08-960-048-1	Sequence 1, Appli
	29	29	0.9	3328	4	US-09-838-586-1	Sequence 1, Appli
	30	29	0.9	4072	3	US-09-245-041-16	Sequence 16, Appl
	31	29	0.9	4612	3	US-08-960-048-2	Sequence 2, Appli
	32	29	0.9	4612	4	US-09-838-586-2	Sequence 2, Appli
	33	29	0.9	6370	3	US-09-245-041-12	Sequence 12, Appl
	34	29	0.9	8589	3	US-09-245-041-14	Sequence 14, Appl
c	35	29	0.9	9646	3	US-08-811-566-1	Sequence 1, Appli
c	36	29	0.9	9646	4	US-09-034-756-1	Sequence 1, Appli
c	37	29	0.9	12980	3	US-08-811-566-5	Sequence 5, Appli
c	38	29	0.9	12980	4	US-09-034-756-5	Sequence 5, Appli
	39	28	0.9	3046	1	US-08-726-725-1	Sequence 1, Appli
	40	27	0.8	184	4	US-09-480-921B-13	Sequence 13, Appl
	41	27	0.8	767	1	US-08-184-604-1	Sequence 1, Appli
	42	27	0.8	1363	4	US-09-280-116-55	Sequence 55, Appl
	43	27	0.8	2178	3	US-09-034-916-1	Sequence 1, Appli
	44	27	0.8	3603	4	US-09-221-013A-5	Sequence 5, Appli
	45	27	0.8	3673	4	US-09-221-013A-11	Sequence 11, Appl

ALIGNMENTS

RESULT 1
US-09-014-416-57/c
; Sequence 57, Application US/09014416
; Patent No. 6153421
; GENERAL INFORMATION:
; APPLICANT: Yanagi, Masayuki
; APPLICANT: Bukh, Jens
; APPLICANT: Emerson, Susanne U.
; APPLICANT: Purcell, Robert H.
; TITLE OF INVENTION: CLONED GENOMES OF INFECTIOUS HEPATITIS C VIRUSES AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: 20264276
; CURRENT APPLICATION NUMBER: US/09/014,416
; CURRENT FILING DATE: 1998-01-27
; EARLIER APPLICATION NUMBER: US 60/053,062
; EARLIER FILING DATE: 1997-07-18
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 57
; LENGTH: 81
; TYPE: DNA
; ORGANISM: Hepatitis C virus
US-09-014-416-57

Query Match 0.9%; Score 30; DB 3; Length 81;
Best Local Similarity 100.0%; Pred. No. 0.0056;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3203 GAAAAAAAAAAAGAAAAAAAAAAAAAAAAA 3232
||||||||||||||||||||||||||||||
Db 41 GAAAAAAAAAAAGAAAAAAAAAAAAAAAAA 12

RESULT 2
US-09-014-416-59/c
; Sequence 59, Application US/09014416
; Patent No. 6153421
; GENERAL INFORMATION:
; APPLICANT: Yanagi, Masayuki
; APPLICANT: Bukh, Jens
; APPLICANT: Emerson, Susanne U.
; APPLICANT: Purcell, Robert H.
; TITLE OF INVENTION: CLONED GENOMES OF INFECTIOUS HEPATITIS C VIRUSES AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: 20264276
; CURRENT APPLICATION NUMBER: US/09/014,416
; CURRENT FILING DATE: 1998-01-27
; EARLIER APPLICATION NUMBER: US 60/053,062
; EARLIER FILING DATE: 1997-07-18
; NUMBER OF SEQ ID NOS: 65

SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 59
LENGTH: 127
TYPE: DNA
ORGANISM: Hepatitis C virus
-09-014-416-59

Query Match 0.9%; Score 30; DB 3; Length 127;
Best Local Similarity 100.0%; Pred. No. 0.0053;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

3203 GAAAAAAAAAAAGAAAAAAAAAAAAAAAAA 3232
84 GAAAAAAAAAAAGAAAAAAAAAAAAAAAAA 55

SULT 3
-09-014-416-63/c
Sequence 63, Application US/09014416
Patent No. 6153421
GENERAL INFORMATION:
APPLICANT: Yanagi, Masayuki
APPLICANT: Bukh, Jens
APPLICANT: Emerson, Susanne U.
APPLICANT: Purcell, Robert H.
TITLE OF INVENTION: CLONED GENOMES OF INFECTIOUS HEPATITIS C VIRUSES AND
TITLE OF INVENTION: USES THEREOF
FILE REFERENCE: 20264276
CURRENT APPLICATION NUMBER: US/09/014,416
CURRENT FILING DATE: 1998-01-27
EARLIER APPLICATION NUMBER: US 60/053,062
EARLIER FILING DATE: 1997-07-18
NUMBER OF SEQ ID NOS: 65
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 63
LENGTH: 176
TYPE: DNA
ORGANISM: Hepatitis C virus
-09-014-416-63

Query Match 0.9%; Score 30; DB 3; Length 176;
Best Local Similarity 100.0%; Pred. No. 0.0051;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

3203 GAAAAAAAAAAAGAAAAAAAAAAAAAAAAA 3232
84 GAAAAAAAAAAAGAAAAAAAAAAAAAAAAA 55

SULT 4
-09-014-416-60/c
Sequence 60, Application US/09014416
Patent No. 6153421
GENERAL INFORMATION:
APPLICANT: Yanagi, Masayuki
APPLICANT: Bukh, Jens
APPLICANT: Emerson, Susanne U.
APPLICANT: Purcell, Robert H.
TITLE OF INVENTION: CLONED GENOMES OF INFECTIOUS HEPATITIS C VIRUSES AND
TITLE OF INVENTION: USES THEREOF
FILE REFERENCE: 20264276
CURRENT APPLICATION NUMBER: US/09/014,416
CURRENT FILING DATE: 1998-01-27
EARLIER APPLICATION NUMBER: US 60/053,062
EARLIER FILING DATE: 1997-07-18
NUMBER OF SEQ ID NOS: 65
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 60
LENGTH: 183
TYPE: DNA
ORGANISM: Hepatitis C virus
-09-014-416-60

Query Match [illegible]; Score [illegible]; DB 3; Length 183;
Best Local Similarity [illegible]; Pred. No. [illegible];
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3203 GAAAAAAAAAAAGAAAAAAAAAAAAAAAAA 3232
Db 84 GAAAAAAAAAAAGAAAAAAAAAAAAAAAAA 55

RESULT 5
US-09-014-416-64/c
; Sequence 64, Application [illegible]
; Patent No. 6153421
; GENERAL INFORMATION:
; APPLICANT: Yanagi, Masay[illegible]
; APPLICANT: Bukh, Jens
; APPLICANT: Emerson, Susa[illegible]
; APPLICANT: Purcell, Robe[illegible]
; TITLE OF INVENTION: [illegible] HEPATITIS C VIRUSES AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: 20264276
; CURRENT APPLICATION NUMBER: US/09/014,416
; CURRENT FILING DATE: 1998-01-27
; EARLIER APPLICATION NUMBER: US 60/053,062
; EARLIER FILING DATE: 1997-07-18
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 64
; LENGTH: 200
; TYPE: DNA
; ORGANISM: Hepatitis C virus
US-09-014-416-64

Query Match [illegible]; Score 30; DB 3; Length 200;
Best Local Similarity 100.0%; Pred. No. 0.0051;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3203 GAAAAAAAAAAAGAAAAAAAAAAAAAAAAA 3232
Db 60 GAAAAAAAAAAAGAAAAAAAAAAAAAAAAA 31

RESULT 6
US-09-014-416-4/c
; Sequence 4, Application US/09014416
; Patent No. 6153421
; GENERAL INFORMATION:
; APPLICANT: Yanagi, Masayuki
; APPLICANT: Bukh, Jens
; APPLICANT: Emerson, Susanne U.
; APPLICANT: Purcell, Robert H.
; TITLE OF INVENTION: CLONED GENOMES OF INFECTIOUS HEPATITIS C VIRUSES AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: 20264276
; CURRENT APPLICATION NUMBER: US/09/014,416
; CURRENT FILING DATE: 1998-01-27
; EARLIER APPLICATION NUMBER: US 60/053,062
; EARLIER FILING DATE: 1997-07-18
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 9595
; TYPE: DNA
; ORGANISM: Hepatitis C virus
US-09-014-416-4

Query Match 0.9%; Score 30; DB 3; Length 9595;
Best Local Similarity 100.0%; Pred. No. 0.0033;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3203 GAAAAAAAAAAAGAAAAAAAAAAAAAAAAA 3232
Db 9454 GAAAAAAAAAAAGAAAAAAAAAAAAAAAAA 9425

```
RESULT 7
US-09-014-416-2/c
; Sequence 2, Application US/09014416
; Patent No. 6153421
; GENERAL INFORMATION:
;  APPLICANT: Yanagi, Masayuki
;  APPLICANT: Bukh, Jens
;  APPLICANT: Emerson, Susanne U.
;  APPLICANT: Purcell, Robert H.
;  TITLE OF INVENTION: CLONED GENOMES OF INFECTIOUS HEPATITIS C VIRUSES AND
;  TITLE OF INVENTION: USES THEREOF
;  FILE REFERENCE: 20264276
;  CURRENT APPLICATION NUMBER: US/09/014,416
;  CURRENT FILING DATE: 1998-01-27
;  EARLIER APPLICATION NUMBER: US 60/054,162
;  EARLIER FILING DATE: 1997-07-18
;  NUMBER OF SEQ ID NOS: 65
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 9599
;   TYPE: DNA
;   ORGANISM: Hepatitis C virus
US-09-014-416-2

  Query Match             0.9%;  Score 30;  DB 3;  Length 9599;
  Best Local Similarity   100.0%;  Pred. No. 0.0033;
  Matches   30;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3203 GAAAAAAAAAAAAGAAAAAAAAAAAAAAAAAA 3232
             |||||||||||||||||||||||||||||||| 
Db       9458 GAAAAAAAAAAAAGAAAAAAAAAAAAAAAAAA 9429


RESULT 8
US-09-014-416-6/c
; Sequence 6, Application US/09014416
; Patent No. 6153421
; GENERAL INFORMATION:
;  APPLICANT: Yanagi, Masayuki
;  APPLICANT: Bukh, Jens
;  APPLICANT: Emerson, Susanne U.
;  APPLICANT: Purcell, Robert H.
;  TITLE OF INVENTION: CLONED GENOMES OF INFECTIOUS HEPATITIS C VIRUSES AND
;  TITLE OF INVENTION: USES THEREOF
;  FILE REFERENCE: 20264276
;  CURRENT APPLICATION NUMBER: US/09/014,416
;  CURRENT FILING DATE: 1998-01-27
;  EARLIER APPLICATION NUMBER: US 60/054,162
;  EARLIER FILING DATE: 1997-07-18
;  NUMBER OF SEQ ID NOS: 65
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
;   LENGTH: 9599
;   TYPE: DNA
;   ORGANISM: Hepatitis C virus
US-09-014-416-6

  Query Match             0.9%;  Score 30;  DB 3;  Length 9599;
  Best Local Similarity   100.0%;  Pred. No. 0.0033;
  Matches   30;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3203 GAAAAAAAAAAAAGAAAAAAAAAAAAAAAAAA 3232
             |||||||||||||||||||||||||||||||| 
Db       9458 GAAAAAAAAAAAAGAAAAAAAAAAAAAAAAAA 9429


RESULT 9
US-08-664-596B-9
; Sequence 9, Application US/08664596B
; Patent No. 5807703
;  GENERAL INFORMATION:
;    APPLICANT:  Jacobs, Kenneth
;    APPLICANT:  McCoy, John
;    APPLICANT:  LaVallie, Edward
;    APPLICANT:  Racie, Lisa
;    APPLICANT:  Merberg, David
;    APPLICANT:  Treacy, Maurice
;    APPLICANT:  Evans, Cheryl
;    APPLICANT:  Spaulding, Vikki
;    APPLICANT:  Bowman, Michael
;    TITLE OF INVENTION:  SECRETED PROTEINS AND POLYNUCLEOTIDES
;    TITLE OF INVENTION:  ENCODING THEM
;    NUMBER OF SEQUENCES:  37
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genetics Institute, Inc.
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/664,596B
;      FILING DATE:
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brown, Scott A.
;      REGISTRATION NUMBER:  32,724
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 498-8224
;      TELEFAX:  (617) 876-5851
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  44 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-664-596B-9

  Query Match             0.9%;  Score 29;  DB 1;  Length 44;
  Best Local Similarity   100.0%;  Pred. No. 0.015;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3204 AAAAAAAAAAAGAAAAAAAAAAAAAAAAA 3232
             |||||||||||||||||||||||||||||
Db          3 AAAAAAAAAAAGAAAAAAAAAAAAAAAAA 31


RESULT 10
US-08-520-678A-22/c
; Sequence 22, Application US/08520678A
; Patent No. 5874565
;  GENERAL INFORMATION:
;    APPLICANT:  Rice, Charles M.
;    APPLICANT:  Kolykhalov, Alexander A.
;    TITLE OF INVENTION:  NOVEL 3' TERMINAL SEQUENCE OF HEPATITIS
;    TITLE OF INVENTION:  C VIRUS GENOME AND DIAGNOSTIC AND THERAPEUTIC USES THEREOF
;    NUMBER OF SEQUENCES:  39
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Howell & Haferkamp, L.C.
;      STREET:  7733 Forsyth Blvd., Suite 1400
;      CITY:  St. Louis
;      STATE:  MO
;      COUNTRY:  USA
;      ZIP:  63105
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
```

SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/520,678A
FILING DATE:
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Henderson, Melodie W.
REGISTRATION NUMBER: 37,848
REFERENCE/DOCKET NUMBER: 6029-6836
TELECOMMUNICATION INFORMATION:
TELEPHONE: 314-727-5188
TELEFAX: 314-727-6092
TELEX:
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 356 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
-08-520-678A-22

Query Match 0.9%; Score 29; DB 2; Length 356;
Best Local Similarity 100.0%; Pred. No. 0.012;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

3204 AAAAAAAAAAAGAAAAAAAAAAAAAAAAA 3232
||||||||||||||||||||||||||||
190 AAAAAAAAAAAGAAAAAAAAAAAAAAAAA 162

SULT 11
-08-897 126-22/c
Sequence 22, Application US/08897126
Patent No. 6297003
GENERAL INFORMATION:
APPLICANT: Rice, Charles M.
APPLICANT: Kolykhalov, Alexander A.
TITLE OF INVENTION: NOVEL 3' TERMINAL SEQUENCE OF HEPATITIS
TITLE OF INVENTION: C VIRUS GENOME AND DIAGNOSTIC AND THERAPEUTIC USES THEREOF
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: Howell & Haferkamp, L.C.
STREET: 7733 Forsyth Blvd., Suite 1400
CITY: St. Louis
STATE: MO
COUNTRY: USA
ZIP: 63105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/897,126
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/520,678
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Henderson, Melodie W.
REGISTRATION NUMBER: 37,848
REFERENCE/DOCKET NUMBER: 6029-6836
TELECOMMUNICATION INFORMATION:
TELEPHONE: 314-727-5188
TELEFAX: 314-727-6092
TELEX:
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 356 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-897-126-22

Query Match [illegible]; Score 29; DB 4; Length 356;
Best Local Similarity [illegible]; Pred. No. 0.012;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3204 AAAAAAAAAAAGAAAAAAAAAAAAAAAAA 3232
||| |||| ||| | | ||| | ||
Db 190 AAAAAAAAAAAGAAAAAAAAAAAAAAAAA 162

RESULT 12
US-09-720-201A-4/c
; Sequence 4, Application US/09/720,201A
; Patent No. 6524853
; GENERAL INFORMATION:
; APPLICANT: KOHARA, MICHIN
; APPLICANT: KOHARA, KYOF
; APPLICANT: TAIRA, KAZUNA
; APPLICANT: MATSUZAKI,
; APPLICANT: ORMORI, HIS
; TITLE OF INVENTION: A VE EXPRESSING AN RNA VIRAL FULL-LENGTH GENE AND
; TITLE OF INVENTION: ITS USE
; FILE REFERENCE: 04853.0051 00000
; CURRENT APPLICATION NUMBER: US/09/720,201A
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: JP 98/177,6
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: PCT/JP99/0
; PRIOR FILING DATE: 1999-06-24
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 537
; TYPE: DNA
; ORGANISM: Hepatitus C Virus
US-09-720-201A 4

Query Match 0.9%; Score 29; DB 4; Length 537;
Best Local Similarity 100.0%; Pred. No. 0.012;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3204 AAAAAAAAAAAGAAAAAAAAAAAAAAAAA 3232
||||:|||||||| ||| |||| | ||
Db 374 AAAAAAAAAAAGAAAAAAAAAAAAAAAAA 346

RESULT 13
US-09-357-251-7
; Sequence 7, Application US/09357251
; Patent No. 6271441
; GENERAL INFORMATION:
; APPLICANT: Falco, S. Carl
; APPLICANT: Famodu, Layo
; APPLICANT: Orozco, Buddy
; APPLICANT: Schwaber, James S.
; TITLE OF INVENTION: Plant Aminoacyl tRNA Synthetase
; FILE REFERENCE: BB-1193
; CURRENT APPLICATION NUMBER: US/09/357,251
; CURRENT FILING DATE: 1999-07-20
; EARLIER APPLICATION NUMBER: 60/093,
; EARLIER FILING DATE: July 21, 1998
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 7
; LENGTH: 572
; TYPE: DNA
; ORGANISM: Triticum aestivum
US-09-357-251-7

Query Match 0.9%; Score 29; DB 3; Length 572;

Best Local Similarity 100.0%; Pred. No. 0.012;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3204 AAAAAAAAAAAAGAAAAAAAAAAAAAAAAA 3232
Db 543 AAAAAAAAAAAAGAAAAAAAAAAAAAAAAA 571

RESULT 14
US-09-328-111-147/c
; Sequence 147, Application US/09328111
; Patent No. 6262333
; GENERAL INFORMATION:
; APPLICANT: Endege, Wilson O.
; APPLICANT: Steinmann, Kathleen E.
; APPLICANT: Astle, Jon H.
; APPLICANT: Burgess, Christopher
; APPLICANT: Bushnell, Steven E.
; APPLICANT: Carroll III, Eddie
; APPLICANT: Catino, Theodore J.
; APPLICANT: Derti, Adnan
; APPLICANT: Ford, Donna M.
; APPLICANT: Lewis, Marcia E.
; APPLICANT: Monahan, John E.
; APPLICANT: Schlegel, Robert
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; TITLE OF INVENTION: PRODUCTS
; FILE REFERENCE: CCD-257 (US)
; CURRENT APPLICATION NUMBER: US/09/328,111
; CURRENT FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: US 60/088,901
; EARLIER FILING DATE: 1998-06-10
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: FastSEQ for Windows Version
; SEQ ID NO 147
; LENGTH: 599
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(599)
; OTHER INFORMATION: n = A,T,C or G
US-09-328-111-147

Query Match 0.9%; Score 29; DB 3; Length 599;
Best Local Similarity 100.0%; Pred. No. 0.012;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3204 AAAAAAAAAAAAGAAAAAAAAAAAAAAAAA 3232
Db 52 AAAAAAAAAAAAGAAAAAAAAAAAAAAAAA 80

RESULT 15
US-09-385-982-293/c
; Sequence 293, Application US/09385982
; Patent No. 6262334
; GENERAL INFORMATION:
; APPLICANT: ENDEGE, WILSON O., ET AL.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; TITLE OF INVENTION: PRODUCTS: II
; FILE REFERENCE: CCDNA-260XX
; CURRENT APPLICATION NUMBER: US/09/385,982
; CURRENT FILING DATE: 1999-08-30
; EARLIER APPLICATION NUMBER: 09/328,111
; EARLIER FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: 60/117,393
; EARLIER FILING DATE: 1999-01-27
; EARLIER APPLICATION NUMBER: 60/098,619
; EARLIER FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 544
; SOFTWARE: FastSEQ for Windows Version
; SEQ ID NO 293
; LENGTH: 603
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(603)
; OTHER INFORMATION: n = A,T,C or G
US-09-385-982-293

Query Match 0.9%; Score 29; DB 3; Length 603;
Best Local Similarity 100.0%; Pred. No. 0.011;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3204 AAAAAAAAAAAAGAAAAAAAAAAAAAAAAA 3232
Db 49 AAAAAAAAAAAAGAAAAAAAAAAAAAAAAA 21

Search completed: October 31, 2003, 09:32:59
Job time : 172.895 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 31, 2003, 09:47:13 ; Search time 904.611 Seconds
(without alignments)
10924.991 Million cell updates/sec

Title: US-09-980-043-1
Perfect score: 3232
Sequence: 1 gtcgacccacgcgtccgtt.....aaaaaaaaaaaaaaaaaaaaa 3232

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1811591 seqs, 1359896190 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3623182

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score listed being printed, and is derived by analysis of the database.

SUMMARIES

Result No.		Score	Query Match	Length	DB	ID	Description
c	1	30	0.9	309	10	US-09-880-107-452	Sequence 452, App
	2	30	0.9	473	11	US-09-918-995-2985	Sequence 2985, Ap
	3	30	0.9	473	14	US-10-108-946-9908	Sequence 9908, Ap
	4	30	0.9	1131	11	US-09-884-878-42	Sequence 42, Appl
	5	30	0.9	1131	11	US-09-885-236-42	Sequence 42, Appl
	6	30	0.9	2667	11	US-09-895-298-12	Sequence 12, Appl
	7	30	0.9	4167	9	US-09-764-878-282	Sequence 282, App
	8	30	0.9	4167	9	US-09-764-878-1145	Sequence 1145, Ap
c	9	30	0.9	4167	10	US-09-764-846-345	Sequence 345, App
c	10	30	0.9	4167	11	US-09-764-891-7710	Sequence 7710, Ap
c	11	30	0.9	4167	11	US-09-764-891-10231	Sequence 10231, A
c	12	30	0.9	4167	14	US-10-091-483-345	Sequence 345, App
	13	30	0.9	4167	14	US-10-079-854-282	Sequence 282, App
	14	30	0.9	4167	14	US-10-074-095-1145	Sequence 1145, Ap
c	15	30	0.9	4167	14	US-10-205-428-1019	Sequence 1019, Ap
c	16	30	0.9	10803	10	US-09-747-419-17	Sequence 17, Appl
c	17	30	0.9	10803	14	US-10-259-275-17	Sequence 17, Appl
	18	29	0.9	64	10	US-09-983-965-42	Sequence 42, Appl
	19	29	0.9	67	10	US-09-983-965-245	Sequence 245, App
	20	29	0.9	76	10	US-09-960-352-6176	Sequence 6176, Ap
c	21	29	0.9	78	14	US-10-060-036-2238	Sequence 2238, Ap
	22	29	0.9	80	10	US-09-764-846-267	Sequence 267, App
	23	29	0.9	80	10	US-09-764-846-339	Sequence 339, App
	24	29	0.9	80	14	US-10-091-483-267	Sequence 267, App
	25	29	0.9	80	14	US-10-091-483-339	Sequence 339, App
c	26	29	0.9	92	13	US-10-029-913-3	Sequence 3, Appli
c	27	29	0.9	98	10	US-09-878-178-1575	Sequence 1575, Ap
c	28	29	0.9	98	13	US-10-046-935-1575	Sequence 1575, Ap
c	29	29	0.9	98	14	US-10-146-502-1575	Sequence 1575, Ap
c	30	29	0.9	101	9	US-09-838-386-23	Sequence 23, Appl
c	31	29	0.9	103	14	US-10-066-543-3251	Sequence 3251, Ap
c	32	29	0.9	107	14	US-10-066-543-2236	Sequence 2236, Ap
c	33	29	0.9	107	14	US-10-102-524-1055	Sequence 1055, Ap
	34	29	0.9	110	11	US-09-746-783-195	Sequence 195, App
c	35	29	0.9	119	14	US-10-102-524-229	Sequence 229, App
c	36	29	0.9	123	14	US-10-102-524-240	Sequence 240, App
c	37	29	0.9	136	10	US-09-960-352-7557	Sequence 7557, Ap
	38	29	0.9	142	10	US-09-878-574-13342	Sequence 13342, A
c	39	29	0.9	143	14	US-10-060-036-2351	Sequence 2351, Ap
	40	29	0.9	145	11	US-09-991-936-925	Sequence 925, App
c	41	29	0.9	147	14	US-10-066-543-2095	Sequence 2095, Ap
	42	29	0.9	148	10	US-09-878-574-5369	Sequence 5369, Ap
c	43	29	0.9	151	10	US-09-764-846-277	Sequence 277, App
c	44	29	0.9	151	10	US-09-764-846-299	Sequence 299, App
c	45	29	0.9	151	10	US-09-764-846-341	Sequence 341, App

ALIGNMENTS

RESULT 1
US-09-880-107-452/c
; Sequence 452, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 452
; LENGTH: 309
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 AA226925
US-09-880-107-452

Query Match 0.9%; Score 30; DB 10; Length 309;
Best Local Similarity 100.0%; Pred. No. 0.00049;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3202 TGAAAAAAAAAAAAGAAAAAAAAAAAAAAAA 3231
||||||||||||||||||||||||||||||
Db 30 TGAAAAAAAAAAAAGAAAAAAAAAAAAAAAA 1

RESULT 2
US-09-918-995-2985
; Sequence 2985, Application US/09918995

```
Publication No. US20030073623A1
GENERAL INFORMATION:
 APPLICANT: Hyseq, Inc.
 TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
 TITLE OF INVENTION:  FROM VARIOUS cDNA LIBRARIES
 FILE REFERENCE: 20411-756
 CURRENT APPLICATION NUMBER: US/09/918,995
 CURRENT FILING DATE:  2001-07-30
 PRIOR APPLICATION NUMBER: US/09/235,076
 PRIOR FILING DATE: 1999-01-20
 NUMBER OF SEQ ID NOS: 38054
 SOFTWARE: FastSEQ for Windows Version 3.0
SEQ ID NO 2985
  LENGTH: 473
  TYPE: DNA
  ORGANISM: Homo sapiens
  FEATURE:
  NAME/KEY: misc_feature
  LOCATION: (1)..(473)
  OTHER INFORMATION: n = A,T,C or G
-09-918-995-2985

Query Match              0.9%;  Score 30;  DB 11;  Length 473;
Best Local Similarity   100.0%;  Pred. No. 0.0005;
Matches   30;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

       3203 GAAAAAAAAAAAGAAAAAAAAAAAAAAAAA 3232
            ||||||||||||||||||||||||||||||
        245 GAAAAAAAAAAAGAAAAAAAAAAAAAAAAA 274


SULT 3
-10-198-846-9908
Sequence 9908, Application US/10198846
Publication No. US20030099974A1
GENERAL INFORMATION:
 APPLICANT: Lillie, James
 APPLICANT:  Xu, Yongyao
 APPLICANT:  Wang, Youzhen
 APPLICANT:  Steinmann, Kathleen
 TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
 TITLE OF INVENTION:  FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
 TITLE OF INVENTION:  THERAPY OF BREAST CANCER
 FILE REFERENCE: MRI-049
 CURRENT APPLICATION NUMBER: US/10/198,846
 CURRENT FILING DATE:  2002-07-18
 PRIOR APPLICATION NUMBER: 60/306,220
 PRIOR FILING DATE: 2001-07-18
 NUMBER OF SEQ ID NOS: 14084
 SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 9908
  LENGTH: 473
  TYPE: DNA
  ORGANISM: Homo sapiens
  FEATURE:
  NAME/KEY: misc_feature
  LOCATION: 1
  OTHER INFORMATION: n = A,T,C or G
-10-198-846-9908

Query Match              0.9%;  Score 30;  DB 14;  Length 473;
Best Local Similarity   100.0%;  Pred. No. 0.0005;
Matches   30;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

       3203 GAAAAAAAAAAAGAAAAAAAAAAAAAAAAA 3232
            ||||||||||||||||||||||||||||||
        188 GAAAAAAAAAAAGAAAAAAAAAAAAAAAAA 217


SULT 4
-09-974-879-42
Sequence 42, Application US/09974879
Publication No. US20030028003A1
```

```
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: 125 Human Secreted Proteins
;  FILE REFERENCE: PZ020P2
;  CURRENT APPLICATION NUMBER: US/09/974,879
;  CURRENT FILING DATE:  2001-10-12
;  PRIOR APPLICATION NUMBER: US 60/239,811
;  PRIOR FILING DATE: 2000-10-13
;  PRIOR APPLICATION NUMBER: US 09/918,995
;  PRIOR FILING DATE: 2001-07-30
;  PRIOR APPLICATION NUMBER: US 09/305,736
;  PRIOR FILING DATE: 1999-05-05
;  PRIOR APPLICATION NUMBER: PCT/US98/23435
;  PRIOR FILING DATE: 1998-11-04
;  PRIOR APPLICATION NUMBER: US 60/064,911
;  PRIOR FILING DATE: 1997-11-07
;  PRIOR APPLICATION NUMBER: US 60/064,912
;  PRIOR FILING DATE: 1997-11-07
;  PRIOR APPLICATION NUMBER: US 60/064,913
;  PRIOR FILING DATE: 1997-11-07
;  PRIOR APPLICATION NUMBER: US 60/064,914
;  PRIOR FILING DATE: 1997-11-07
;  PRIOR APPLICATION NUMBER: US 60/064,915
;  PRIOR FILING DATE: 1997-11-07
;  PRIOR APPLICATION NUMBER: US 60/064,916
;  PRIOR FILING DATE: 1997-11-07
;  PRIOR APPLICATION NUMBER: US 60/064,917
;  PRIOR FILING DATE: 1997-11-07
;  PRIOR APPLICATION NUMBER: US 60/064,918
;  PRIOR FILING DATE: 1997-11-07
;  PRIOR APPLICATION NUMBER: US 60/064,919
;  PRIOR FILING DATE: 1997-11-07
;  PRIOR APPLICATION NUMBER: US 60/064,920
;  PRIOR FILING DATE: 1997-11-07
;  PRIOR APPLICATION NUMBER: US 60/064,921
;  PRIOR FILING DATE: 1997-11-07
;  PRIOR APPLICATION NUMBER: US 60/064,922
;  PRIOR FILING DATE: 1997-11-07
;  NUMBER OF SEQ ID NOS: 611
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 42
;   LENGTH: 1131
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-974-879-42

  Query Match              0.9%;  Score 30;  DB 11;  Length 1131;
  Best Local Similarity   100.0%;  Pred. No. 0.00051;
  Matches   30;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3203 GAAAAAAAAAAAGAAAAAAAAAAAAAAAAA 3232
             ||||| |||||.||, ||| |||..'||
Db       1070 GAAAAAAAAAAAGAAAAAAAAAAAAAAAAA 1099


RESULT 5
US-09-305-736-42
; Sequence 42, Application US/09305736
; Publication No. US20030087968A1
; GENERAL INFORMATION:
;  APPLICANT: Feng et al.
;  TITLE OF INVENTION: 125 Human Secreted Proteins
;  FILE REFERENCE: PZ020P1
;  CURRENT APPLICATION NUMBER: US/09/305,736
;  CURRENT FILING DATE: 1999-05-05
;  EARLIER APPLICATION NUMBER: PCT/US98/23435
;  EARLIER FILING DATE: 1998-11-04
;  EARLIER APPLICATION NUMBER: 60/064,911
;  EARLIER FILING DATE: 1997-11-07
```

```
;  EARLIER APPLICATION NUMBER: 60/064,912
;  EARLIER FILING DATE: 1997-11-07
;  EARLIER APPLICATION NUMBER: 60/064,983
;  EARLIER FILING DATE: 1997-11-07
;  EARLIER APPLICATION NUMBER: 60/064,900
;  EARLIER FILING DATE: 1997-11-07
;  EARLIER APPLICATION NUMBER: 60/064,988
;  EARLIER FILING DATE: 1997-11-07
;  EARLIER APPLICATION NUMBER:  60/064,987
;  EARLIER FILING DATE:1997-11-07
;  EARLIER APPLICATION NUMBER: 60/064,908
;  EARLIER FILING DATE: 1997-11-07
;  EARLIER APPLICATION NUMBER: 60/064,984
;  EARLIER FILING DATE: 1997-11-07
;  EARLIER APPLICATION NUMBER:  60/064,985
;  EARLIER FILING DATE: 1997-11-07
;  EARLIER APPLICATION NUMBER: 60/066,094
;  EARLIER FILING DATE: 1997-11-17
;  EARLIER APPLICATION NUMBER: 60/066,100
;  EARLIER FILING DATE: 1997-11-17
;  EARLIER APPLICATION NUMBER: 60/066,089
;  EARLIER FILING DATE: 1997-11-17
;  EARLIER APPLICATION NUMBER: 60/066,095
;  EARLIER FILING DATE: 1997 11-17
;  EARLIER APPLICATION NUMBER: 60/066,090
;  EARLIER FILING DATE: 1997 11-17
;  NUMBER OF SEQ ID NOS: 612
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 42
;    LENGTH: 1131
;    TYPE: DNA
;    ORGANISM: Homo sapiens
US-09-305 736-42

  Query Match             0.9%;  Score 30;  DB 11;  Length 1131;
  Best Local Similarity   100.0%;  Pred. No. 0.00051;
  Matches   30;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3203 GAAAAAAAAAAAGAAAAAAAAAAAAAAAAA 3232
             ||||||||||||||||||||||||||||||
Db       1070 GAAAAAAAAAAAGAAAAAAAAAAAAAAAAA 1099


RESULT 6
US 09-895-298-12
; Sequence 12, Application US/09895298
; Publication No. US20030078405A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: 47 Human Secreted Proteins
;  FILE REFERENCE: PZ035P1
;  CURRENT APPLICATION NUMBER: US/09/895,298
;  CURRENT FILING DATE:  2001-07-02
;  PRIOR APPLICATION NUMBER: 09/591,16
;  PRIOR FILING DATE: 2000-06-09
;  PRIOR APPLICATION NUMBER: PCT/US99/29950
;  PRIOR FILING DATE: 1999-12-16
;  PRIOR APPLICATION NUMBER: 60/113,056
;  PRIOR FILING DATE: 1998-12-18
;  PRIOR APPLICATION NUMBER: 60/112,869
;  PRIOR FILING DATE: 1998-12-17
;  NUMBER OF SEQ ID NOS: 231
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
;    LENGTH: 2667
;    TYPE: DNA
;    ORGANISM: Homo sapiens
US-09 895-298-12

  Query Match             0.9%;  Score 30;  DB 11;  Length 2667;
  Best Local Similarity   100.0%;  Pred. No. 0.00052;
  Matches   30;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3203 GAAAAAAAAAAAGAAAAAAAAAAAAAAAAA 3232
             ||||||||||||||||||||||||||||||
Db       2579 GAAAAAAAAAAAGAAAAAAAAAAAAAAAAA 2608


RESULT 7
US-09-764-878-282
; Sequence 282, Application US/09764878
; Patent No. US20020090615A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PA121
;  CURRENT APPLICATION NUMBER: US/09/764,878
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 428
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 282
;    LENGTH: 4167
;    TYPE: DNA
;    ORGANISM: Homo sapiens
;    FEATURE:
;    NAME/KEY: SITE
;    LOCATION: (2901)
;    OTHER INFORMATION: n equals a,t,g, or c
;    NAME/KEY: SITE
;    LOCATION: (2902)
;    OTHER INFORMATION: n equals a,t,g, or c
;    NAME/KEY: SITE
;    LOCATION: (2903)
;    OTHER INFORMATION: n equals a,t,g, or c
;    NAME/KEY: SITE
;    LOCATION: (2904)
;    OTHER INFORMATION: n equals a,t,g, or c
;    NAME/KEY: SITE
;    LOCATION: (2905)
;    OTHER INFORMATION: n equals a,t,g, or c
;    NAME/KEY: SITE
;    LOCATION: (2906)
;    OTHER INFORMATION: n equals a,t,g, or c
;    NAME/KEY: SITE
;    LOCATION: (2907)
;    OTHER INFORMATION: n equals a,t,g, or c
;    NAME/KEY: SITE
;    LOCATION: (2908)
;    OTHER INFORMATION: n equals a,t,g, or c
;    NAME/KEY: SITE
;    LOCATION: (2909)
;    OTHER INFORMATION: n equals a,t,g, or c
;    NAME/KEY: SITE
;    LOCATION: (2910)
;    OTHER INFORMATION: n equals a,t,g, or c
;    NAME/KEY: SITE
;    LOCATION: (2911)
;    OTHER INFORMATION: n equals a,t,g, or c
;    NAME/KEY: SITE
;    LOCATION: (2912)
;    OTHER INFORMATION: n equals a,t,g, or c
;    NAME/KEY: SITE
;    LOCATION: (2913)
;    OTHER INFORMATION: n equals a,t,g, or c
;    NAME/KEY: SITE
;    LOCATION: (2914)
;    OTHER INFORMATION: n equals a,t,g, or c
;    NAME/KEY: SITE
;    LOCATION: (2915)
;    OTHER INFORMATION: n equals a,t,g, or c
;    NAME/KEY: SITE
;    LOCATION: (2916)
;    OTHER INFORMATION: n equals a,t,g, or c
;    NAME/KEY: SITE
;    LOCATION: (2917)
```

OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (2918)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (2919)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (2920)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (2921)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (2922)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (2923)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (2924)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (2925)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (2926)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (2927)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (2928)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (2929)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (2930)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (2931)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (2932)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (2933)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (2934)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (2935)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (2936)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (2937)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (2938)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (2939)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (2940)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (2941)
OTHER INFORMATION: n equals a,t,g, or c

; NAME/KEY: SITE
; LOCATION: (2942)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (2943)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (2944)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (2945)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (2946)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (2947)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (2948)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (2949)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (2950)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (2951)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (2952)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (2953)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (2954)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (2955)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (2956)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (2957)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (2958)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (2959)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (2960)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (2961)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE

Query Match 0.4%; Score 30; DB 9; Length 4167;
Best Local Similarity 100.0%; Pred. No. 7.8e-05;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3203 GAAAAAAAAAAAGAAAAAAAAAAAAAAAAA 3232
||||||||||||||:|||||||||:||||
Db 4015 GAAAAAAAAAAAGAAAAAAAAAAAAAAAAA 4044

RESULT 8
US-09-764-860-1145

```
; Sequence 1145, Application US/09764860
; Patent No. US20020094953A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PC008
;  CURRENT APPLICATION NUMBER: US/09/764,860
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 1198
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1145
;   LENGTH: 4167
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (2901)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (2902)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (2903)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (2904)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (2905)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (2906)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (2907)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (2908)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (2909)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (2910)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (2911)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (2912)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (2913)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (2914)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (2915)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (2916)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (2917)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (2918)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (2919)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (2920)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (2921)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (2922)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (2923)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (2924)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (2925)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (2926)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (2927)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (2928)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (2929)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (2930)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (2931)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (2932)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (2933)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (2934)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (2935)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (2936)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (2937)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (2938)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (2939)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (2940)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (2941)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (2942)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (2943)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
```

```
  LOCATION: (2944)
  OTHER INFORMATION: n equals a,t,g, or c
  NAME/KEY: SITE
  LOCATION: (2945)
  OTHER INFORMATION: n equals a,t,g, or c
  NAME/KEY: SITE
  LOCATION: (2946)
  OTHER INFORMATION: n equals a,t,g, or c
  NAME/KEY: SITE
  LOCATION: (2947)
  OTHER INFORMATION: n equals a,t,g, or c
  NAME/KEY: SITE
  LOCATION: (2948)
  OTHER INFORMATION: n equals a,t,g, or c
  NAME/KEY: SITE
  LOCATION: (2949)
  OTHER INFORMATION: n equals a,t,g, or c
  NAME/KEY: SITE
  LOCATION: (2950)
  OTHER INFORMATION: n equals a,t,g, or c
  NAME/KEY: SITE
  LOCATION: (2951)
  OTHER INFORMATION: n equals a,t,g, or c
  NAME/KEY: SITE
  LOCATION: (2952)
  OTHER INFORMATION: n equals a,t,g, or c
  NAME/KEY: SITE
  LOCATION: (2953)
  OTHER INFORMATION: n equals a,t,g, or c
  NAME/KEY: SITE
  LOCATION: (2954)
  OTHER INFORMATION: n equals a,t,g, or c
  NAME/KEY: SITE
  LOCATION: (2955)
  OTHER INFORMATION: n equals a,t,g, or c
  NAME/KEY: SITE
  LOCATION: (2956)
  OTHER INFORMATION: n equals a,t,g, or c
  NAME/KEY: SITE
  LOCATION: (2957)
  OTHER INFORMATION: n equals a,t,g, or c
  NAME/KEY: SITE
  LOCATION: (2958)
  OTHER INFORMATION: n equals a,t,g, or c
  NAME/KEY: SITE
  LOCATION: (2959)
  OTHER INFORMATION: n equals a,t,g, or c
  NAME/KEY: SITE
  LOCATION: (2960)
  OTHER INFORMATION: n equals a,t,g, or c
  NAME/KEY: SITE
  LOCATION: (2961)
  OTHER INFORMATION: n equals a,t,g, or c
  NAME/KEY: SITE

Query Match              0.9%;  Score 30;  DB 9;  Length 4167;
Best Local Similarity   100.0%;  Pred. No. 0.00053;
Matches   30;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

        3203 GAAAAAAAAAAAGAAAAAAAAAAAAAAAAA 3232
             ||||||||||||||||||||||||||||||
>       4015 GAAAAAAAAAAAGAAAAAAAAAAAAAAAAA 4044


ESULT 9
S-09-764-846-345/c
 Sequence 345, Application US/09764846
 Patent No. US20020102638A1
 GENERAL INFORMATION:
  APPLICANT: Rosen et al.
  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
  FILE REFERENCE: PTZ12
  CURRENT APPLICATION NUMBER: US/09/764,846
;  CURRENT FILING DATE:  2 [illegible] 01-[illegible]
;  Prior application data to be used  [illegible] consult PAIM or file wrapper
;  NUMBER OF SEQ ID NOS: 345
;  SOFTWARE: PatentIn Ver  2 [illegible]
; SEQ ID NO 345
;   LENGTH: 4167
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (268)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (269)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (270)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (271)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (272)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (273)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (274)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (275)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (276)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (277)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (278)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (279)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (280)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (281)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (282)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (283)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (284)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (285)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (286)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (287)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (288)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
```

; LOCATION: (289)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (290)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (291)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (292)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (293)
; OTHER INFORMATION: n equals a,t,g, or
; NAME/KEY: SITE
; LOCATION: (294)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (295)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (296)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (297)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (298)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (299)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (300)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (301)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (302)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (303)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (304)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (305)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (306)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (307)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (308)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (309)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (310)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (311)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (312)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (313)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (314)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (315)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (316)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (317)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (318)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (319)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (320)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (321)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (322)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (323)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (324)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (325)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (326)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (327)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (328)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE

Query Match 0.9%; Score 30; DB 10; Length 4167;
Best Local Similarity 100.0%; Pred. No. 0.00053;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       3203 GAAAAAAAAAAAGAAAAAAAAAAAAAAAAA 3232
             ||||||||||||||||||||||||||||||
Db        153 GAAAAAAAAAAAGAAAAAAAAAAAAAAAAA 124
```

RESULT 10
US-09-764-891-7710/c
; Sequence 7710, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7710
; LENGTH: 4167
; TYPE: DNA

ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (268)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (269)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (270)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (271)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (272)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (273)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (274)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (275)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (276)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (277)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (278)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (279)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (280)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (281)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (282)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (283)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (284)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (285)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (286)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (287)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (288)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (289)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (290)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (291)

; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (292)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (293)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (294)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (295)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (296)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (297)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (298)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (299)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (300)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (301)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (302)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (303)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (304)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (305)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (306)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (307)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (308)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (309)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (310)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (311)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (312)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (313)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (314)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (315)
; OTHER INFORMATION: n equals a,t,g, or c

```
;   NAME/KEY: SITE
;   LOCATION: (316)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (317)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (318)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (319)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (320)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (321)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (322)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (323)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (324)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (325)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (326)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (327)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (328)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE

  Query Match              0.9%;  Score 50;  DB 11;  Length 4167;
  Best Local Similarity   100.0%;  Pred. No. 0.00025;
  Matches   30;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        320 GAAAAAAAAAAAGAAAAAAAAAAAAAAAAA 349
              ||||| ||||||||||||||| |||| ||||
Db        153 GAAAAAAAAAAAGAAAAAAAAAAAAAAAAA 124


RESULT 11
US-09-764-691-10231/c
; Sequence 10231, Application US/09/764,691
; Publication No. US20030077806A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PC006
;  CURRENT APPLICATION NUMBER: US/09/764,691
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - C result (ALM or file wrapper
;  NUMBER OF SEQ ID NOS: 10231
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10231
;   LENGTH: 4167
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (268)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (269)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (270)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (271)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (272)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (273)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (274)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (275)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (276)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (277)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (278)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (279)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (280)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (281)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (282)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (283)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (284)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (285)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (286)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (287)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (288)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (289)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (290)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (291)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (292)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (293)
;   OTHER INFORMATION: n equals a,t,g, or c
```

NAME/KEY: SITE
LOCATION: (294)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (295)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (296)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (297)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (298)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (299)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (300)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (301)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (302)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (303)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (304)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (305)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (306)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (307)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (308)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (309)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (310)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (311)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (312)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (313)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (314)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (315)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (316)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (317)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE

; LOCATION: (318)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (319)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (320)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (321)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (322)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (323)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (324)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (325)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (326)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (327)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (328)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE

Query Match [illegible]; Score [illegible]; DB [illegible]; Length 4167;
Best Local Similarity [illegible]; Pred No. [illegible];
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3203 GAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3232
||||||||||||||||||||||||||||||
Db 153 GAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 124

RESULT 12
US-10-091-483-345/c
; Sequence 345, Application US/10091483
; Publication No. US20030049[illegible]A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT2120
; CURRENT APPLICATION NUMBER: US/10/091,483
; CURRENT FILING DATE: [illegible]
; NUMBER OF SEQ ID NOS: 349
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. [illegible]
; SEQ ID NO 345
; LENGTH: 4167
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (268)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (269)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (270)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (271)
; OTHER INFORMATION: n equals a,t,g, or c

; NAME/KEY: misc_feature
; LOCATION: (272)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (273)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (274)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (275)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (276)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (277)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (278)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (279)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (280)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (281)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (282)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (283)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (284)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (285)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (286)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (287)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (288)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (289)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (290)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (291)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (292)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (293)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (294)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (295)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (296)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (297)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (298)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (299)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (300)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (301)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (302)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (303)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (304)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (305)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (306)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (307)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (308)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (309)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (310)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (311)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (312)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (313)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (314)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (315)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (316)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (317)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (318)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (319)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (320)

OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (321)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (322)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (323)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (324)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (325)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (326)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (327)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (328)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature

Query Match 0.9%; Score 30; DB 14; Length 4167;
Best Local Similarity 100.0%; Pred. No. 0.00053;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

3203 GAAAAAAAAAAAGAAAAAAAAAAAAAAAAA 3232
||||||||||||||||||||||||||||||
153 GAAAAAAAAAAAGAAAAAAAAAAAAAAAAA 124

SULT 13
-10-079-854-282
Sequence 282, Application US/10079854
Publication No. US20030054368A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PA121C1
CURRENT APPLICATION NUMBER: US/10/079,854
CURRENT FILING DATE: 2002-02-22
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 428
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 282
LENGTH: 4167
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (2901)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (2902)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (2903)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (2904)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (2905)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (2906)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
; LOCATION: (2907)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (2908)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (2909)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (2910)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (2911)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (2912)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (2913)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (2914)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (2915)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (2916)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (2917)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (2918)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (2919)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (2920)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (2921)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (2922)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (2923)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (2924)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (2925)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (2926)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (2927)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (2928)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (2929)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (2930)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (2931)

; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (2932)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (2933)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (2934)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (2935)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (2936)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (2937)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (2938)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (2939)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (2940)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (2941)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (2942)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (2943)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (2944)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (2945)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (2946)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (2947)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (2948)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (2949)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (2950)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (2951)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (2952)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (2953)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (2954)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (2955)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (2956)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (2957)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (2958)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (2959)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (2960)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (2961)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature

Query Match 0.9%; Score 30; DB 14; Length 4167;
Best Local Similarity 100.0%; Pred. No. 0.00053;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3203 GAAAAAAAAAAAGAAAAAAAAAAAAAAAAA 3232
||||||||||||||||||||||||||||||
Db 4015 GAAAAAAAAAAAGAAAAAAAAAAAAAAAAA 4044

RESULT 14
US-10-074-095-1145
; Sequence 1145, Application US/10074095
; Publication No. US20030077704A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC008C1
; CURRENT APPLICATION NUMBER: US/10/074,095
; CURRENT FILING DATE: 2002-02-14
; PRIOR APPLICATION NUMBER: 09/764,860
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: 60/214,886
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 60/217,487
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,758
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/220,963
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/217,496
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,447
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/218,290
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/225,757
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/226,868
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: 60/216,647
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: 60/225,267
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/216,880
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: 60/225,270
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/251,869
; PRIOR FILING DATE: 2000-12-08

PRIOR APPLICATION NUMBER: 60/235,834
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: 60/234,274
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: 60/234,223
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: 60/228,924
PRIOR FILING DATE: 2000-08-30
PRIOR APPLICATION NUMBER: 60/224,518
PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/236,369
PRIOR FILING DATE: 2000-09-29
PRIOR APPLICATION NUMBER: 60/224,519
PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/220,964
PRIOR FILING DATE: 2000-07-26
PRIOR APPLICATION NUMBER: 60/241,809
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/249,299
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/236,327
PRIOR FILING DATE: 2000-09-29
PRIOR APPLICATION NUMBER: 60/241,785
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/244,617
PRIOR FILING DATE: 2000-11-01
PRIOR APPLICATION NUMBER: 60/225,268
PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/236,368
PRIOR FILING DATE: 2000-09-29
PRIOR APPLICATION NUMBER: 60/251,856
PRIOR FILING DATE: 2000-12-08
PRIOR APPLICATION NUMBER: 60/251,868
PRIOR FILING DATE: 2000-12-08
PRIOR APPLICATION NUMBER: 60/229,344
PRIOR FILING DATE: 2000-09-01
PRIOR APPLICATION NUMBER: 60/234,997
PRIOR FILING DATE: 2000-09-25
PRIOR APPLICATION NUMBER: 60/229,343
PRIOR FILING DATE: 2000-09-01
PRIOR APPLICATION NUMBER: 60/229,345
PRIOR FILING DATE: 2000-09-01
PRIOR APPLICATION NUMBER: 60/229,287
PRIOR FILING DATE: 2000-09-01
PRIOR APPLICATION NUMBER: 60/229,513
PRIOR FILING DATE: 2000-09-05
PRIOR APPLICATION NUMBER: 60/231,413
PRIOR FILING DATE: 2000-09-08
PRIOR APPLICATION NUMBER: 60/229,509
PRIOR FILING DATE: 2000-09-05
PRIOR APPLICATION NUMBER: 60/236,367
PRIOR FILING DATE: 2000-09-29
PRIOR APPLICATION NUMBER: 60/237,039
PRIOR FILING DATE: 2000-10-02
PRIOR APPLICATION NUMBER: 60/237,038
PRIOR FILING DATE: 2000-10-02
PRIOR APPLICATION NUMBER: 60/236,370
PRIOR FILING DATE: 2000-09-29
PRIOR APPLICATION NUMBER: 60/236,802
PRIOR FILING DATE: 2000-10-02
PRIOR APPLICATION NUMBER: 60/237,037
PRIOR FILING DATE: 2000-10-02
PRIOR APPLICATION NUMBER: 60/237,040
PRIOR FILING DATE: 2000-10-02
PRIOR APPLICATION NUMBER: 60/240,960
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/239,935
PRIOR FILING DATE: 2000-10-13
PRIOR APPLICATION NUMBER: 60/239,937
PRIOR FILING DATE: 2000-10-13
PRIOR APPLICATION NUMBER: 60/241,787
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/246,474

; PRIOR FILING DATE: 2000-11-06
; PRIOR APPLICATION NUMBER: 60/246,933
; PRIOR FILING DATE: 2000-11-06
; PRIOR APPLICATION NUMBER: 60/249,211
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,207
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/226,681
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: 60/225,351
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/225,214
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/227,187
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: 60/225,214
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/235,836
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/230,478
; PRIOR FILING DATE: 2000-09-07
; PRIOR APPLICATION NUMBER: 60/225,214
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/249,206
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,208
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,215
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,112
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,193
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,248
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,244
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,217
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,211
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,215
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,264
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,214
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,297
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/232,400
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/231,242
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/232,041
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/232,087
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/231,414
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/231,244
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/233,064
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/233,063
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,397
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,399
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,401
; PRIOR FILING DATE: 2000-09-14

```
;  PRIOR APPLICATION NUMBER: 60/241,808
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/241,826
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/241,746
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/241,221
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/246,475
;  PRIOR FILING DATE: 2000-11-08
;  PRIOR APPLICATION NUMBER: 60/231,241
;  PRIOR FILING DATE: 2000-09-08

  Query Match             0.9%;  Score 40;  DB 14;  Length 4167;
  Best Local Similarity   100.0%;  Pred. No. 5.3e-03;
  Matches   30;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3203 GAAAAAAAAAAAAGAAAAAAAAAAAAAAAAA 3232
             ||||| ||||| ||||| ||||| ||||| |||
Db       4015 GAAAAAAAAAAAAGAAAAAAAAAAAAAAAAA 4044


RESULT 15
US-10-205-428-1019/c
; Sequence 1019, Application US/10205428
; Publication No. US20030108907A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, And Antibodies
;  FILE REFERENCE: PA117C1
;  CURRENT APPLICATION NUMBER: US/10/205,428
;  CURRENT FILING DATE:  2002-07-26
;  PRIOR APPLICATION NUMBER: 09/764,842
;  PRIOR FILING DATE: 2001-01-17
;  PRIOR APPLICATION NUMBER: 60/179,265
;  PRIOR FILING DATE: 2000-01-31
;  PRIOR APPLICATION NUMBER: 60/181,629
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: 60/214,890
;  PRIOR FILING DATE: 2000-06-28
;  PRIOR APPLICATION NUMBER: 60/217,487
;  PRIOR FILING DATE: 2000-07-11
;  PRIOR APPLICATION NUMBER: 60/225,758
;  PRIOR FILING DATE: 2000-08-14
;  PRIOR APPLICATION NUMBER: 60/218,961
;  PRIOR FILING DATE: 2000-07-26
;  PRIOR APPLICATION NUMBER: 60/218,490
;  PRIOR FILING DATE: 2000-07-11
;  PRIOR APPLICATION NUMBER: 60/2[illegible]
;  PRIOR FILING DATE: 2000-07-14
;  PRIOR APPLICATION NUMBER: 60/218,279
;  PRIOR FILING DATE: 2000-07-14
;  Remaining Prior Application data removed - see file wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 1019
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1019
;  LENGTH: 4167
;  TYPE: DNA
;  ORGANISM: Homo sapiens
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: (268)
;  OTHER INFORMATION: n equals a,t,g, or c
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: (269)
;  OTHER INFORMATION: n equals a,t,g, or c
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: (270)
;  OTHER INFORMATION: n equals a,t,g, or c
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: (271)
;  OTHER INFORMATION: n equals a,t,g, or c
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: (272)
;  OTHER INFORMATION: n equals a,t,g, or c
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: (273)
;  OTHER INFORMATION: n equals a,t,g, or c
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: (274)
;  OTHER INFORMATION: n equals a,t,g, or c
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: (275)
;  OTHER INFORMATION: n equals a,t,g, or c
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: (276)
;  OTHER INFORMATION: n equals a,t,g, or c
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: (277)
;  OTHER INFORMATION: n equals a,t,g, or c
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: (278)
;  OTHER INFORMATION: n equals a,t,g, or c
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: (279)
;  OTHER INFORMATION: n equals a,t,g, or c
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: (280)
;  OTHER INFORMATION: n equals a,t,g, or c
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: (281)
;  OTHER INFORMATION: n equals a,t,g, or c
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: (282)
;  OTHER INFORMATION: n equals a,t,g, or c
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: (283)
;  OTHER INFORMATION: n equals a,t,g, or c
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: (284)
;  OTHER INFORMATION: n equals a,t,g, or c
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: (285)
;  OTHER INFORMATION: n equals a,t,g, or c
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: (286)
;  OTHER INFORMATION: n equals a,t,g, or c
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: (287)
;  OTHER INFORMATION: n equals a,t,g, or c
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: (288)
;  OTHER INFORMATION: n equals a,t,g, or c
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: (289)
```

OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (290)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (291)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (292)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (293)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (294)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (295)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (296)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (297)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (298)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (299)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (300)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (301)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (302)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (303)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (304)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (305)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (306)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (307)
OTHER INFORMATION: n equals a,t,g, or c

; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (308)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:

Query Match [illegible] 18.34; Length 4167;
Best Local Similarity [illegible] from N[illegible]3;
Matches 30; Conservative [illegible] Mi[illegible] 0; Indels 0; Gaps 0;

Qy 3203 GAAAAAAAAAAAA[illegible]AAAAAAAAAAAAAAA 3217
|| ||| ||| | | || |
Db 192 GAAAAAAAAAAAA[illegible]AAAAAAAAAAAAAAA 164

Search completed: [illegible]
Job time : 807.611 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 30, 2003, 06:37:05 ; Search time 6342.97 Seconds
(without alignments)
12394.116 Million cell updates/sec

Title: US-09-980-043-1
Perfect score: 3232
Sequence: 1 qtcgaccacgcgtccgtct.....aaGuuu.aaaaaaaaaaaaa 3232

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1133.4	35.1	3898	11	AY110415	AY110415 Zea mays
2	1133	34.1	3897	11	AY103655	AY103655 Zea mays
3	1102.8	34.1	3763	11	AY108113	AY108113 Zea mays
4	1101.8	34.1	3783	11	AY104740	AY104740 Zea mays
5	1093.4	33.8	2872	11	AY104236	AY104236 Zea mays
6	1086.8	33.6	3728	11	AY112236	AY112236 Zea mays
7	1071.2	33.1	3788	11	AY103701	AY103701 Zea mays
8	1065.8	33.0	3696	11	AY110079	AY110079 Zea mays
9	556.6	17.2	571	13	BU895617	BU895617 X028D07 P
10	481.2	14.9	731	13	BQ867361	BQ867361 QGD10J04.
11	475.2	14.7	648	13	BU894292	BU894292 X005H01 P
12	470.6	14.6	800	14	CB648832	CB648832 OSJNEb12G
13	456.8	14.1	484	12	BI128887	BI128887 G083P04Y
14	455.6	14.1	735	13	BQ969585	BQ969585 QHB38P07.
15	453.6	14.0	668	10	BF635224	BF635224 NF061H04D
16	452.8	14.0	1189	11	AY107656	AY107656 Zea mays
17	443.8	13.7	658	10	BE322406	BE322406 NF020G01I
18	440.8	13.6	479	12	BI132326	BI132326 G135P19Y
19	436.6	13.5	792	14	CB671068	CB671068 OSJNEe04H
20	432.8	13.4	688	13	BU099300	BU099300 WHE3305_H
21	431.8	13.4	953	14	CD439980	CD439980 EL01N0531
22	428.2	13.2	457	12	BI131891	BI131891 G127P18Y
23	427.4	13.2	613	13	BU082548	BU082548 saq34c07.
24	426.4	13.2	753	14	CA784277	CA784277 sat96g06.
25	424.6	13.1	799	14	CB894714	CB894714 EST647506
26	421	13.0	476	13	BU897587	BU897587 X060B01 P
27	418.2	12.9	779	14	CA916873	CA916873 EST641020
28	416.2	12.9	573	12	BI784558	BI784558 saf56d07.
29	415.8	12.9	435	12	BI127228	BI127228 G057P27Y
30	415.4	12.9	851	14	CB892280	CB892280 EST649249
31	415	12.8	463	12	BI131187	BI131187 G116P69Y
32	404.2	12.5	783	14	CB671503	CB671503 OSJNEe05B
33	402.4	12.5	620	9	AW309128	AW309128 sf94d11.y
34	402.2	12.4	683	13	BU897715	BU897715 X067B10 P
35	401.4	12.4	625	12	BI309194	BI309194 EST530604
36	401.2	12.4	871	10	BF253907	BF253907 HVSMEf000
c 37	400.6	12.4	607	13	BQ514409	BQ514409 EST621824
38	399.2	12.4	856	14	CB633016	CB633016 OSIIEb11J
39	396.2	12.3	434	13	BU894130	BU894130 X003C01 P
40	395.4	12.2	669	14	CD227355	CD227355 CCC1_50_D
41	393.4	12.2	787	14	CD486470	CD486470 CRH2.4E12
42	392.2	12.1	717	13	BQ996712	BQ996712 QGG13J11.
43	388.6	12.0	619	13	BQ468970	BQ468970 HM02P11r
44	388.6	12.0	714	14	CB629928	CB629928 OSIIEb06I
45	387.6	12.0	978	10	BG321258	BG321258 Zm04_03f1

ALIGNMENTS

RESULT 1
AY110415
LOCUS AY110415 3898 bp mRNA linear HTC 17-OCT 2002
DEFINITION Zea mays CL1166_1 mRNA sequence.
ACCESSION AY110415
VERSION AY110415.1 GI:21214824
KEYWORDS HTC.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 3898)
AUTHORS Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S., Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
TITLE Maize Mapping Project/DuPont Consensus Sequences for Design of Overgo Probes
JOURNAL Unpublished (2002)
REFERENCE 2 (bases 1 to 3898)
AUTHORS Coe,E.H.
TITLE Direct Submission
JOURNAL Submitted (25-APR-2002) Maize Mapping Project, University of Missouri, Columbia, MO 65211, USA
COMMENT If you are interested in getting corresponding physical clones, these are publicly available from ZmDB and may be found by BLAST searching at MSL, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR, www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the

```
          maize cDNA sequences is either Virginia Walbot, Stanford or Pat
          Schnable, Iowa State, then clones may be requested from ZmDB:
          www.zmdb.iastate.edu.
ATURES               Location/Qualifiers
   source            1. .3898
                     /organism="Zea mays"
                     /mol_type="mRNA"
                     /db_xref="MaizeDB:630057"
                     /db_xref="taxon:4577"
                     /clone_lib="Maize Mapping Project/DuPont Cornsensus
                     Library"
                     /note="this sequence is part of a project of EST
                     assemblies resulting from the application of public
                     contigs to seed DuPont contigs; this resource was
                     assembled by DuPont as part of a collaboration for the
                     overgo addressing of BACs in conjunction with the Maize
                     Mapping Project"
SE COUNT      932 a    841 c    991 g   1054 t     80 others
IGIN

Query Match             35.1%;  Score 1133.4;  DB 11;  Length 3898;
Best Local Similarity   68.2%;  Pred. No. 2.6e-147;
Matches 1638;  Conservative    0;  Mismatches  748;  Indels   17;  Gaps    7;

        558 CCAATTCCACGCAACAAGCTCACACCATACAGAGCAGTGATCATTATGCGACTGGTCATT 617
            |||||| |    ||| ||||||  |  ||| | | |||||||||| | || ||  ||||
        969 CCAATTTCCTCAAACCAGCTCAACCTTTACCGGGTAGTGATCATTCTCCGTCTTATCATC 1028

        618 CTGGGCCTCTTCTTCTTCCACTTCAGAATAACAAATCCTGTCGATAGTGCCTTTGGCCTGTGG 677
            ||| || |||||||||||  |   | || |   |||| ||    | ||| | |||  |||:|
       1029 CTGTGCTTCTTCTTCTTCCAGTATCGTATCAGTCATCCAGTGCGTGATGCTTATGGATTGTGG 1088

        678 CTTACTTCTGTCATATGTGAGATCTGGTTTGCATTTTCTTGGGTGTTGGATCAGTTCCCC 737
            ||    ||||| || || ||| |||||||||| || || ||| |  | ||:||||||||
       1089 CTAGTATCTGTTATCTGCGAGGTCTGGTTTGCCTTGTCCTGGCTTCTAGATCAGTTCCCA 1148

        738 AAGTGGAATCCTGTCAATAGAGAAACGTATATCGAAAGGCTGTCGGCAAGGTATGAAAGA 797
            || ||: |||| |||| | || || ||| |||| |||||  |    |||||||| |||
       1149 AAATGGTATCCAATCAACCGTGAGACATATCTCGACAGGCTTGCATTGAGGTATGATAGA 1208

        798 GAGGGTGAGCCTTCTCAGCTTGCTGGTGTGGATTTTTTCGTGAGTACTGTTGATCCGCTG 857
            ||||| ||||| || ||||| |||    | ||| | ||||| |||||  | |||||  ||
       1209 GAGGGAGAGCCATCACAGCTGGCTCCCATTGATGTCTTCGTCAGTACAGTGGATCCATTG 1268

        858 AAGGAACCGCCATTGATCACTGCCAATACAGTCCTTTCCATCCTTGCTGTGGACTATCCC 917
            |||||||| ||| |||||||| ||||| || ||  | ||||| ||||||||||| || ||
       1269 AAGGAACCTCCACTGATCACAGCCAACACTGTTTTGTCCATTCTTGCTGTGGATTACCCT 1328

        918 GTCGATAAAGTCTCCTGCTACGTGTCTGATGATGGTGCAGCTATGCTTTCATTTGAATCT 977
            || || ||||| || |||| || |||||||||||  |||:|||| |  | ||||| |||
       1329 GTTGACAAAGTGTCATGCTATGTTTCTGATGATGGCTCAGCTATGCTGACTTTTGAGTCT 1388

        978 CTTGTAGAAACAGCTGAGTTTGCAAGGAAGTGGGTTCCGTTCTGCAAAAAATTCTCAATT 1037
            ||    ||||| || || |||||| || ||||||||||| |: || || ||    |  ||
       1389 CTCTCTGAAACCGCCGAATTTGCTAGAAAGTGGGTTCCCTTTTGTAAGAAGCACAATATT 1448

       1038 GAACCAAGAGCACCGGAGTTTTACTTCTCACAGAAAATTGATTACTTGAAAGACAAGGTT 1097
            |||||||||| || || |||||||  | |      | ||||||  |:|| |||||  ||
       1449 GAACCAAGAGCTCCAGAATTTTACTTTGCTCNNNNNNTAGATTACCTGAAGGACAAAATT 1508

       1098 CAACCTTCTTTCGTGAAAGAACGTAGAGCAATGAAAAGGGATTATGAAGAGTACAAAGTC 1157
            :|||||| || || || ||| |  | |||||||| ||||| ||||||||  | ||||||
       1509 CAACCTTCATTTGTTAAGGAAAGACGCGCAATGAAGAGGGAGTATGAAGAATTCAAAGTA 1568

       1158 CGAGTTAATGCCCTGGTAGCAAAGGCTCAGAAAACACCTGAAGAAGGATGGACTATGCAA 1217
             || | |||||||| || || || || |||:||   |||||||| ,| |||||  |||
       1569 AGAATCAATGCCCTTGTTGCCAAAGCACAGAAAGTGCCTGAAGAGGGGTGGACCATGGCT 1628

       1218 GATGGAACACCTTGGCCTGGGAATAACACACGTGATCACCCTGGGCATGATTCAGGTCTT 1277
            |||||||| ||||||||||||||||  |  || || ||| |||||||||||||||  ||
       1629 GATGGAACTGCTTGGCCTGGGAATAACCCTAGGGACCATCCT-GGCATGATTCAGGTGTT 1687

       1278 CCTTGGGAAATACTGGGAGCTCGTGACATTGAAGGAAATGAACTACCTCGTCTAGTATAT 1337
```

```
Db     1688 CTTGGGGCACAGTG GGGCT  ACATTGATGGAAATGAGTTACCAGGTCTTGTCTAT 1745

Qy     1338 GTCTCCAGGGAGAA AGACCTGGCTA CAGCACACAAAAAGGCTGGTGCAGAAAATGCT 1397
            |||| | | |:  ||| |||  |||| |||| |||||||||||  | ||
Db     1746 GTCTCTCGTGAAAA AGACCAGGCTT CAGCATCACAAGAAGGCTGGTGCAATGAATGTG 1805

Qy     1398 CTGGTGAGAGTGTC  CAGTACT AAAATGTCCCTACATCCTCAATGTTGATTGTGAT 1457
            || | | || :| | ||  :||  ||   |||  ||| ||| | |||
Db     1806 CTGATTGTGTATT  GTGTGCTGACAAATGGTGCCTATCTTCTCAATGTGGATTGCGAC 1865

Qy     1458 CACTATGTAAACAA AGCAAG TTGTTCAGAGGCAATGTGCATCCTGATGGACCCACAA 1517
            || '|  ,|| '  ||:, || || ,| |'|.|||| || | .|||| .|
Db     1866 CATTACTTCAATAC AGCAAA TGTTTAGAGAAGCAATGTGCTTCATGATGGATCCGGCT 1925

Qy     1518 GTAGGTCGAGATTC TGCTATGTTCA TTCTCTCAGAGGTTTGATGGCATAGATAAGAGT 1577
            ||,, |  |  :  ||  ||  |  |||  |||.'|||'  ||   .
Db     1926 CTAGGAAAGAAAA  TTTATTA AATTTT ACAGAGATTTGATGGCATTCACTTGCAC 1985

Qy     1578 GATTCTATTGCTAT  TAA CTAT  CTATTTAACATGAAA GCTTGGATGGC 1637
            |||  | |  ,  ||  |  | ,  '|||  ||  : ||||.
Db     1986 GATTGACATTTAA  AAACATA TTT  TTT ATATCAACATGAAAGGTCTGGATGGC 2045

Qy     1638 ATTCAAGGACCAGTATACGTAGGAA TGTTTTGTTTCAACAGGCAAGCACTTTACGGC 1697
            |||| ,  |||  |||| |  |  ||  ||||  |  |  ||
Db     2046 ATTCAGGGTCCAGT TAGGTGGGAA AGTATTGTTTCAATAGGCAGGCTTTGTATGGA 2105

Qy     1698 TACGGGCCTCTCTT   CTATGCCCAGCTTA GCAAGAGAAAGGATTCTTCATCCTGTTT 1757
            |||| ||  ||     ||   ||  |  |  |  |  |
Db     2106 TACGATCCTGTTTT ACTGAAGCTGATCT GAGCCTAACATTGTTGTTAAGAGCTGCNNN 2165

Qy     1758 CTCATGTTGCTGCC TCAAAGAAGAAG TTCAAGATCCAGCTGAGGTATACAGAGA 1813
                                                ||   ||  :  |.|  |||.|
Db     2166 NNNNNNNNNNNNN NNNNNNNNNNNNATAT GATAGTCAAAGCCGTATTATGAAGAGA 2225

Qy     1814 TGCAAAAAGAGAGG TCCAATGCT GATATTTAATCTTACAGAGATTGATAATTATGA 1873
            .|| |    | |  |                 |  |||  .  |  |  | |
Db     2226 -ACAGAATTTTAG TCCCATCTTCAACATGGAAGACATCGAGGAGGGTATTGAAGGTTA 2284

Qy     1874 CGAGCATGAAAGGT AATGCTGAT TGCCAGTTGAGCTTTGAGAAAACTTTTGGCTTATC 1933
            ||,|  ||:|||| | |:| |  || ||||||  |  ||||  ||
Db     2285 CGAGGATGAAAGGT AGTGCTTATGCCCCAGAGGAAATTGGAGAAACGCTTTGGTCAGTC 2344

Qy     1934 TTCTGTCTTCATTGA TCTACACTAATGTACAATGGAGGAGTACCCGAGTCTGCCAATCC 1993
            | |  ||||.||||  ||  |  |  ,|  | || |  ||||     |  | ||  |
Db     2345 TCCAATCTTCATTG ATCCAGCTTTATGACTTAAGGTGGCATACCACCTTCAACAAACCC 2404

Qy     1994 ACCACCATTCATCAAGGAAGCGATTCAAGTCATCGGCTGTGGCTATGAAGAGAAGACTGA 2053
            | |  |  | |  |  |:|||  ||  | | |||| |||| |  | ||  |
Db     2405 AGCTTCTCTACTAAATGAAGCTATCCATGTCATCAGTTGTGGATATGAGGACAAAACTGA 2464

Qy     2054 ATGGGAAAAACAGAT GGTTGGATATATGTGTCAGTCACTGAGGATATCTTAAGTGGCTT 2113
            |||||||||| :| |  ||  | ||  | | | || ||||||  |  | ||| :
Db     2465 ATGGGAAAAGAGA TGGCTGAACTATGTTTCAGTTACGGAGGATATTCTGACTGGCTT 2524

Qy     2114 CAAGATGCACTGC AAGATGAAGATCAATTTACTGCATGCCCGTAAGGCCTGCATTCAA 2173
            .|  |||  |||  |||  |||  |||||  |  | |||  |||||
Db     2525 TAAANNNNNNNNA TGGCTGGTAATCAATCTACTGCATGCCACCACGACCTTGTTTCAA 2584

Qy     2174 AGGATCTGCACCCA AACCTTCTCTATATTTACCAGGTCCTCCGATGGGCTCTTGG 2233
            || ||  |||  |  |  |||  |  |  |||  | |||  | |||| ||||
Db     2585 GGGTTCTGCACCAA AATCTTTCTGATCTTCAATCAGGTGCTCCGTTGGGCTCTTGG 2644

Qy     2234 TTCTGTGGAAATTC GTTTAGCAGACATCTTCTGGTACGGGTTTGGAGGAGGTTG 2293
            || |||||  |||  |||||||||  ||| |||||| || |  |||  |
Db     2645 GTCAGTGGAAATTC GCTTAGCAGACATCTTCTATCTGGTATGGTTACAATGGA-- GG 2701

Qy     2294 TCTTAAATGGCTCCAAAGGCTTGCGTATATAAACACCATTGTGTACCCATTTACATCCCT 2353
            | ,|      |  ||||  ||  | |  ||||||| || ,|| | :|||  |
Db     2702 ATTGAAGCTTTTGGAGAGGCTGGCTTACATCAACACCATTGTTTATCCAATCACATCTGT 2761

Qy     2354 CCCTCTCATTGCCTATTGCACAATTCCTGCAGTTTGTCTGCTCACCGGAAAATTCATCAT 2413
            || || || ||||,||     |||:|||    |||||  || |||   ||||| |||||
```

```
Db       2762 TCCGCTTATCGCCTATTGTGTGCTTCCTGCTATCTGTCTTCTTACCAATAAATTTATCAT 2821

Qy       2414 ACCAACGCTCTCAAACCTGGCAAGCATGCTGTTTCTTGGCCTCTTTATCTCCATCATTGT 2473

Db       2822 TCCTGAGATTAGCAATTATGCTGGGATGTTCTTCATTCTTCTTTTCGCCTCCANNNNGC 2881

Qy       2474 AACTGCGGTGCTTGAGCTAAGATGGAGCGGTGTCAGCATTGAAGATTTATGGCGTAATGA 2533

Db       2882 CACTGGTATATTGGAGCTTAGATGGAGTGGTGTTGGCATTGAAGATTGGTGGAGAAATGA 2941

Qy       2534 ACAATTCTGGGTGATCGGAGGTGTTTCAGTCGATCTCTTTCGGTCTTCCAGGGATTCTT 2593

Db       2942 GCAGTTTTGGGTTATTGGTGGCACCTCTGCCATCTCTTCTAGTGTTCCAGGGTCTGCT 3001

Qy       2594 AAAAATGTTGGCTGGCATCGATACGAACTTCACTGTCACAGTAAAAGCA---GCCGAAGA 2650

Db       3002 GAAAGTGTTGGCTGGGATTGATACCAACTTCACAGTTACCTCAAAGGCATCTGATGAGGA 3061

Qy       2651 TGCAGAATTTGGGGAGCTATATATGGTCAAGTGGACAACACTTTTGATTCCTCCAACCAC 2710

Db       3062 TGGCGACTTTGCTGAGCTATATGTGTTCAAGTGGACCAGTTTGCTCATTCCTCCGACCAC 3121

Qy       2711 ACTTCTCATTATCAATATGTCGGGG-TTGTGCTGGATTCTCTGATGCACTCAACAAAGG 2767

Db       3122 TGTTCTTGTCATTAACCTGGTCGGAATGGTGGCAGGAATTTCGTATGCCATTAACAGTGG 3181

Qy       2768 ATATGAAGCATGGGGGCCTCTCTTTGGCAAGGTGTTCTTTGTTTCTGGGTGATTCTTCA 2827

Db       3182 CTACCAATCCTGGGGTCCGCTCTTTGGAAAGCTGTTCTTCTCGATCTGGGTGNTCCTCCA 3241

Qy       2828 TCTCTATCCATTCCTTAAAGGTCTAATGGTTCGCCAAAACCTAACACCAACCATTGTTGT 2887

Db       3242 TCTGTACACGTTCCTCAAGGGTCTCATGGGAAGGCAGAACCGCACACCAACAATCGTCAT 3301

Qy       2888 TCTCTGGTCAGTGCTGTTGGCCTCTGTCTTCTCTCTCGTTTGGGTCAAGATCAATCCATT 2947

Db       3302 TGTCTGGTCCATCCTTCTTGCATCTATCTTCTCCTTGCTGTGGGTGAAGATCGATCCTTT 3361

Qy       2948 CGT 2950

Db       3362 CAT 3364
```

RESULT 2
AY103655
LOCUS AY103655 [illegible] bp mRNA [illegible] HTC 16-OCT-2002
DEFINITION Zea mays PCO096398 mRNA sequence.
ACCESSION AY103655
VERSION AY103655.1 GI:21206733
KEYWORDS HTC.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 3897)
AUTHORS Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S., Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
TITLE Maize Mapping Project/DuPont Consensus Sequences for Design of Overgo Probes
JOURNAL Unpublished (2002)
REFERENCE 2 (bases 1 to 3897)
AUTHORS Coe,E.H.
TITLE Direct Submission
JOURNAL Submitted (25-APR-2002) Maize Mapping Project, University of Missouri, Columbia, MO 65211, USA
COMMENT If you are interested in getting corresponding physical clones, these are publicly available from ZmDB and may be found by BLAST searching at MSU, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR, www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov When the source of the maize cDNA sequences is either Virginia Walbot, Stanford or Pat Schnable, Iowa State, then clones may be requested from ZmDB: www.zmdb.iastate.edu.

```
FEATURES             Location/Qualifiers
     source          1. .3897
                     /organism="Zea mays"
                     /mol_type="mRNA"
                     /db_xref="MaizeDB:635496"
                     /db_xref="taxon:4577"
                     /clone_lib="Maize Mapping Project/DuPont Consensus Library"
                     /note="this sequence is part of a project of EST assemblies resulting from the application of public contigs to seed DuPont contigs; this resource was assembled by DuPont as part of a collaboration for the overgo addressing of BACs in conjunction with the Maize Mapping Project"
BASE COUNT      982 a    839 c   1008 g   1068 t
ORIGIN
```

Query Match 34.1%; Score 1103; DB 11; Length 3897;
Best Local Similarity 66.1%; Pred. No. 4.1e-143;
Matches 1741; Conservative 0; Mismatches 835; Indels 57; Gaps 8;

```
Qy        385 ATGAATATGGGAATCCAATTTGGAAGAATCGGGTGAAGAGCTGTAAGGATAAAGAGAACA 444

Db        784 ATGGGTATGGTAGTGTTGCTTGGAAGGAACGGATGGAGAATTGGAAGCAGAGACAAGAGA 843

Qy        445 AGAAGAAAAAGAGAAGTCCTAAGGCTGAAACTGAACCAGCTCAAGTTCCTACAGAACAGC 504

Db        844 GGATGCACCAGACGGGGAATGATGGTGGTGGTGATGATGGTGACGATGCTGATCTACCAC 903

Qy        505 AGATGGAAGAGAAACCGTCTGCAGAGGCTTCGGAGCCGCTTTCAATTGTTTATCCAATTC 564

Db        904 TAATGGATGAA--------------GCAAGACAACAACTGTCCAGGAAAATTCCACTTC 948

Qy        565 CACGCAACAAGCTCACACCATACAGAGCAGTGATCATTATGCGACTGGTCATTCTGGGCC 624

Db        949 CATCAAGCCAGATTAATCCATATAGGATGATTATCATTATTCGGCTTGTGGTTTTGGGGT 1008

Qy        625 TCTTCTTCCACTTCAGAATAACAAATCCTGTCGATAGTGCCTTTGGCCTGTGGCTTACTT 684

Db       1009 TCTTCTTCCACTACCGAGTGATGCATCCGGTGAATGATGCATTTGCTTTGTGGCGCATAT 1068

Qy        685 CTGTCATATGTGAGATCTGGTTTGCATTTTCTTGGGTGTTGGATCAGTTCCCCAAGTGGA 744

Db       1069 CTGTTATCTGTGAAATCTGGTTTGCCATGTCTTGGATTCTTGATCAATTCCCAAAGTGGT 1128

Qy        745 ATCCTGTCAATAGAGAAACGTATATCGAAAGGCTGTCGGCAAGGTATGAAAGAGAGGGTG 804

Db       1129 TCCCTATTGAGAGAGAGACTTACCTAGACCGGCTGTCACTGAGGTTCGACAAGGAAGGCC 1188

Qy        805 AGCCTTCTCAGCTTGCTGGTGTGGATTTTTTCGTGAGTACTGTTGATCCGCTGAAGGAAC 864

Db       1189 AGCCATCTCAACTTGCTCCAATTGATTTCTTTGTCAGTACGGTTGATCCCTTAAAGGAAC 1248

Qy        865 CGCCATTGATCACTGCCAATACAGTCCTTTCCATCCTTGCTGTGGACTATCCCGTCGATA 924

Db       1249 CTCCTTTGGTCACAACAAATACTGTTCTATCTATCCTTTCGGTGGATTATCCTGTTGATA 1308

Qy        925 AAGTCTCCTGCTACGTGTCTGATGATGGTGCAGCTATGCTTTCATTTGAATCTCTTGTAG 984

Db       1309 AGGTTTCTTGCTATGTTTCTGATGATGGTGCTGCAATGCTAACGTTTGAAGCATTATCTG 1368

Qy        985 AAACAGCTGAGTTTGCAAGGAAGTGGGTTCCGTTCTGCAAAAAATTCTCAATTGAACCAA 1044

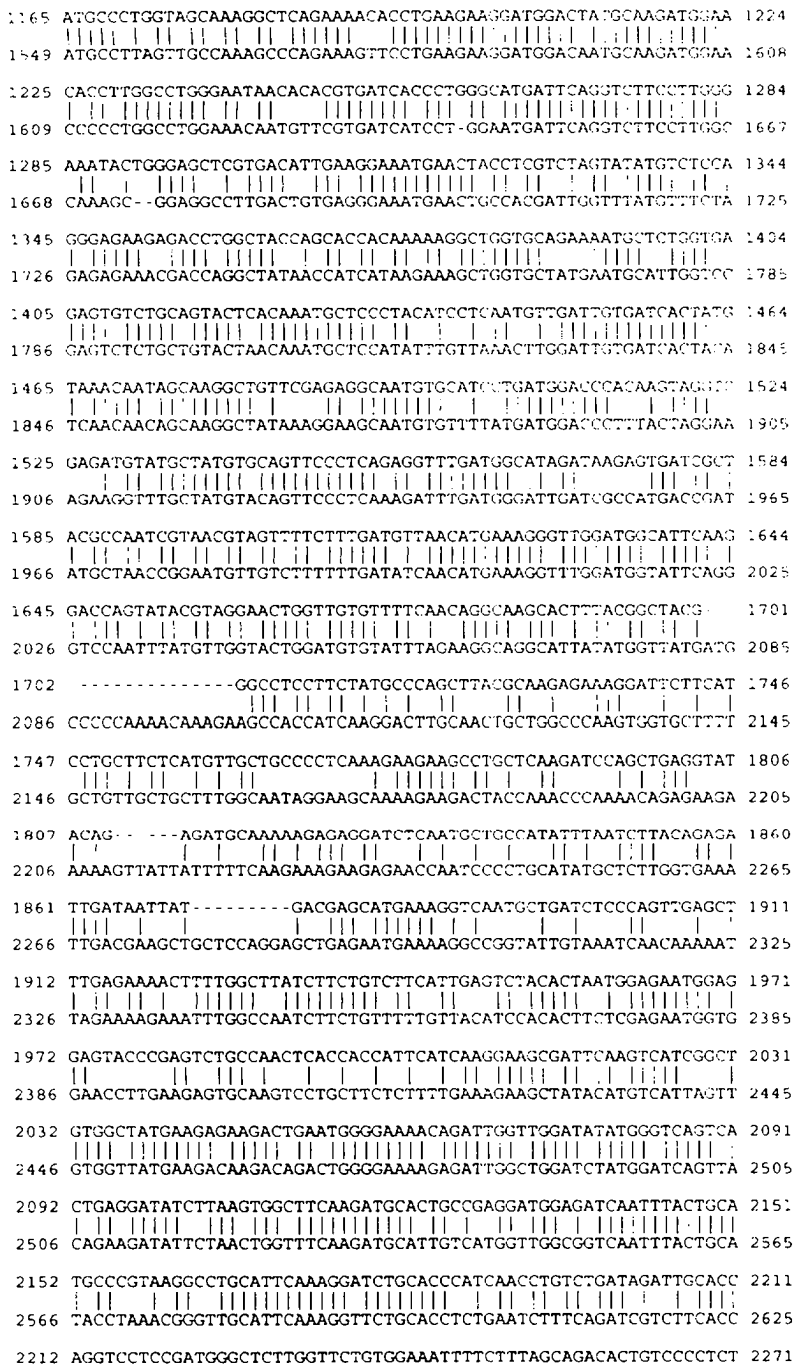
Db       1369 AAACATCTGAATTTGCAAAGAAATGGGTTCCTTTCTGCAAACGGTACAATATTGAACCTC 1428

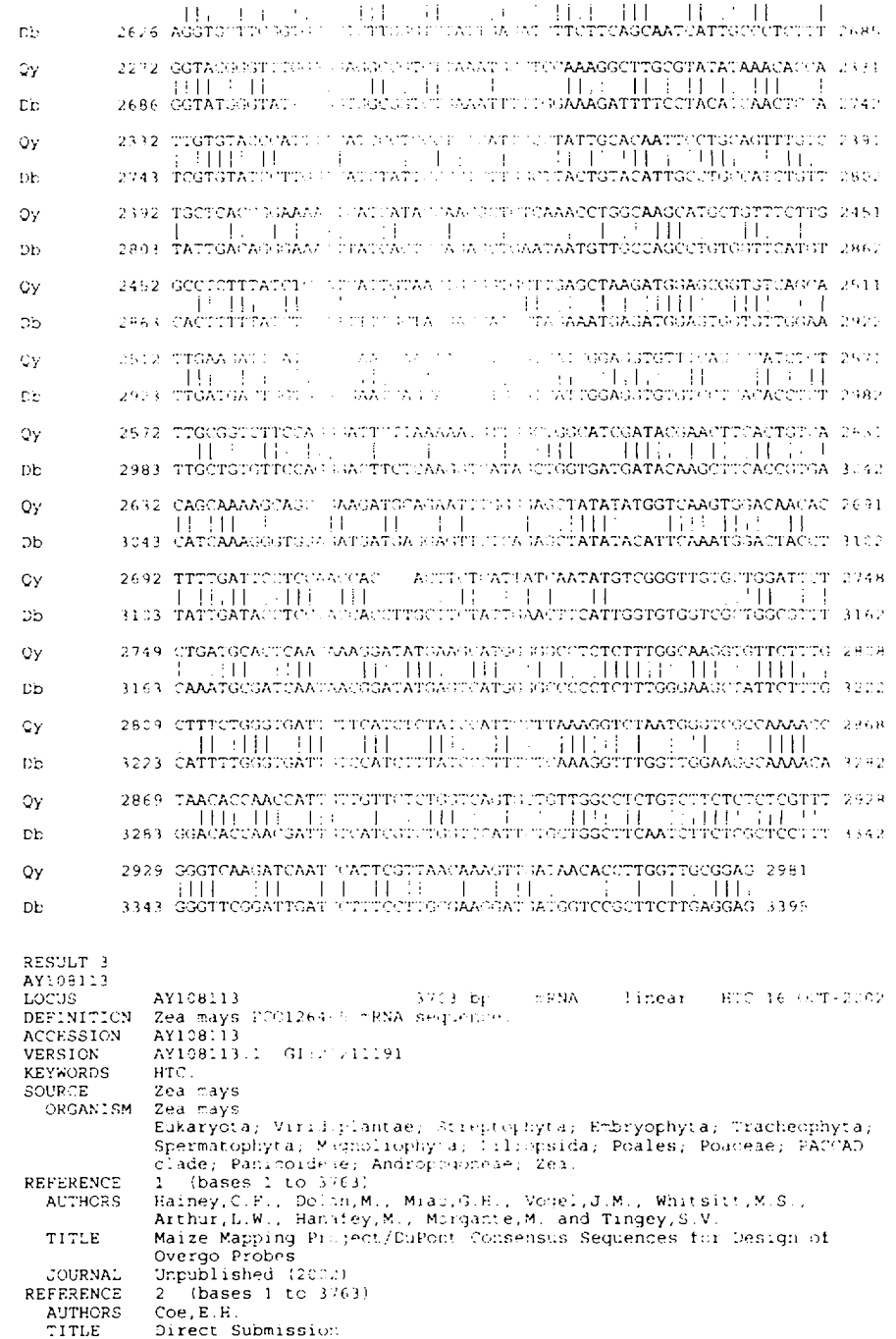
Qy       1045 GAGCACCGGAGTTTTACTTCTCACAGAAAATTGATTACTTGAAAGACAAGGTTCAACCTT 1104

Db       1429 GCGCTCCAGAGTGGTACTTCCAACAGAAGATAGACTACTTGAAAGACAAGGTGGCAGCAA 1488

Qy       1105 CTTTCGTGAAAGAACGTAGAGCAATGAAAAGGGATTATGAAGAGTACAAAGTCCGAGTTA 1164

Db       1489 ACTTTGTTAGGGAGAGGAGAGCAATGAAGAGAGAGTATGAGGAATTCAAGGTGAGAATCA 1548
```





RESULT 3
AY108113
LOCUS AY108113 3763 bp mRNA linear HTC 16-OCT-2002
DEFINITION Zea mays PCO126[illegible] mRNA sequence.
ACCESSION AY108113
VERSION AY108113.1 GI:[illegible]11191
KEYWORDS HTC.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 3763)
AUTHORS Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S., Arthur,L.W., Hannafey,M., Morgante,M. and Tingey,S.V.
TITLE Maize Mapping Project/DuPont Consensus Sequences for Design of Oligo Probes
JOURNAL Unpublished (2002)
REFERENCE 2 (bases 1 to 3763)
AUTHORS Coe,E.H.
TITLE Direct Submission

```
  JOURNAL   Submitted (25-APR-2002) Maize Mapping Project, University of
            Missouri, Columbia, MO 65211, USA
COMMENT     If you are interested in getting corresponding physical clones,
            these are publicly available from ZmDB and may be found by BLAST
            searching at MSL, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR,
            www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the
            maize cDNA sequences is either Virginia Walbot, Stanford or Pat
            Schnable, Iowa State, then clones may be requested from ZmDB:
            www.zmdb.iastate.edu.
FEATURES             Location/Qualifiers
     source          1..3763
                     /organism="Zea mays"
                     /mol_type="mRNA"
                     /db_xref="MaizeDB:637463"
                     /db_xref="taxon:4577"
                     /clone_lib="Maize Mapping Project/DuPont Cornsensus
                     Library"
                     /note="this sequence is part of a project of EST
                     assemblies resulting from the application of public
                     contigs to seed DuPont contigs; this resource was
                     assembled by DuPont as part of a collaboration for the
                     overgo addressing of BACs in conjunction with the Maize
                     Mapping Project"
BASE COUNT      863 a    984 c   1011 g    904 t       others
ORIGIN

  Query Match           34.1%;  Score 1102.8;  DB 1;  Length 3763;
  Best Local Similarity   67.5%;  Pred. No. 4.4e-143;
  Matches 1641;  Conservative    1;  Mismatches  773;  Indels   15;  Gaps    6;

Qy        928 GAGGCTTCGGAGCCGCTTTCAATTGTTTATCCAATTCCAGCAACAAGCTCACACCATAC 987
              |:  ||    |:|| || | |       ..|| |:||   |  '' | |  . |':|||
Db       1097 GAAACTCGACAGCCTCTATCTAGCAAAGTTCCACTTCCTCTTCCAGGATAAATCCATAC 1156

Qy        988 AGAGCAGTGATCATTATGCGACTGGTCATTCTTGTGCGCCTCTTTGCACTTCAGAATAACA 647
              ..   || |:  |:|||| :' |  :||     || :|| | || | | ||| |||
Db       1157 AGGATGGTCATTGTGCTGCGATTGATTGTTCTAAGTATCTTCTTGCACTACCGTATCACA 1216

Qy        648 AATCCTGTCGATAGTGCCTTTGGCCTTGTGGCTTACTTCTTCTATATGTGAGATCTGGTTT 707
              |':|||:   | | | |   | |:'||   |||  |||:'|||| ||||'|||
Db       1217 AATCCTGTGCGCAATGCATACGCATTATGCTTCTATCTTCATATGTGAGATCTGGTTT 1276

Qy        708 GCATTTTCTTGGGTGTTGGATCAGTTCCCCAAGTGGAATCTGTCAATAGAGAAACGTAT 767
              |:  |||' ||| | ||||':|||| | |||||   |:   |||'  |':|| ||
Db       1277 GCTCTTTCGTGGATATTGGATCAGTTCCCTAAGTGGTTTCAATCAACCGGGAGACGTAC 1336

Qy        768 ATCGAAAGGCTGTCGGCAAGGTATCAAAAGCAAGTCGAT   GTCAGCTTGTTGTGTGTG 827
              |    ||:| | |  ||||  :|    | |   ||:  | |  ||| | |:|
Db       1337 CTTGATAGGCTGGCATTAAGTATGAGCTCAAAT  GCT  TTAATTGCTGCTGTT 1396

Qy        828 GATTCTTTGTGAGTACTGTTGATCGTTAAATAAATC   TGATCACTGTCAAATACA 887
              ||  :.'|||' ||||| || |':|    ||||  | |  : |||| :|||||||
Db       1397 GACATTTTCGTCAGTACAGTCGAGCCAATAAAAGTAGCCTC  TTGTCACTGCCAATACC 1456

Qy        888 GTCCTTTCCATCCTTGCTGTGGACTATGCCGTCGATAAAG  TCCTGCTACGTGTCTGAT 947
              || || ||||| |||:|||||  | || :  || ||:|| |  || ||||| |: ||:||
Db       1457 GTGCTATCCATTCTTGCTGTGGATTACCCTCTGGATAAGA  TCTTGCTATGTATCTGAT 1516

Qy        948 GATGGTGCAGCTATGCTTTCATTTGAATCTCTGTAGAAAGAGCTGAGTTTGCAAGGAAG 1007
              |':'; || || |||||  ||||':  |  :  |  ||:  |  |  |'':|||: || ||
Db       1517 GATGGAGCTGCGATGCTGACATTTGATGTACTAGCTGAGACTCAGAGTTTGCTAGAAAA 1576

Qy       1008 TGGGTTCCGTTCTGCAAAAAATTCTCAATTGAACCAAGAGTACCGGAGTTTTACTTCTCA 1067
              |||:| || ||    :| || : |   :||||' | | :|| .: || |  ||||||:
Db       1577 TGGGTACCATTTGTTAAGAAGTACAACATTGAACCTAGACTCCTGAATGGTACTTCTCC 1636

Qy       1068 CAGAAAATTGATTACTTGAAAGACAAGGTTCAATCTTCTTCTGTGAAAGAACGTAGAGCA 1127
              || ||||:||||:|||||, ||||| || :  ||| : || :  ||: | |'  | :
Db       1637 CAGAAAATTGATTACTTGAAGGACAAATTTCACGCTTCGATTCTTAAAGACCGCCGGGCC 1696

Qy       1128 ATGAAAAGGGATTATGAAGAGTACAAAAGTCGAGTTAATGCTTGGTAGCAAAGGCTCAG 1187
              '|.|| |. || ||'|||| | ||: ||  :  ||  ||  :  || | ''|| ' |
Db       1697 ATGAAGAGAGAATATGAAGAATTCAAATTTAGGTAAATGTTTTGTTGCTAAGGACAG 1756

Qy       1188 AAAACACCTGAAGAAGGATGGACTATGCAAGATGGAACACCTTGGCCTGGGAATAACACA 1247
              |||   |||| |||||| |||  ||||||||||| ||||| ||||| || || || ||
Db       1757 AAAGTTCCTGAGGAAGGATGGATCATGCAAGATGGCACACCATGGCCAGGAAACAATACC 1816

Qy       1248 CGTGATCACCCTGGGCATGATTCAGGTCTTCCTTGGGAAATACTGGGAGCTCGTGACATT 1307
              :| || || ||| || || |||||||| |||||||  |     || |  | ||| | |
Db       1817 MGGGACCATCCT-GGAATGATTCAGGTTTTCCTTGGTCA--CAGTGGTGGCCTTGATACT 1873

Qy       1308 GAAGGAAATGAACTACCTCGTCTAGTATATGTCTCCAGGGAGAAGAGACCTGGCTACCAG 1367
              || || |||| ||||  || | || |||| ||  | || ||| | |||| | ||||
Db       1874 GAGGGCAATGAGCTACCCCGTTTGGTCTATGTTTCTCGTGAAAAGCGTCCTGGATTCCAG 1933

Qy       1368 CACCACAAAAAGGCTGGTGCAGAAAATGCTCTGGTGAGAGTGTCTGCAGTACTCACAAAT 1427
              || |||| || |||| |    |||||||| || | || || || || || || |||
Db       1934 CATCACAAGAAAGCTGGTGCCATGAATGCTCTTGTTCGTGTCTCAGCTGTGCTTACCAAT 1993

Qy       1428 GCTCCCTACATCCTCAATGTTGATTGTGATCACTATGTAAACAATAGCAAGGCTGTTCGA 1487
              |  | |||||  | ||| ||||||||||||||||  | ||||| || |||||| |  |
Db       1994 GGACAATACATGTTGAATCTTGATTGTGATCACTACATTAACAACAGTAAGGCTCTCAGG 2053

Qy       1488 GAGGCAATGTGCATCCTGATGGACCCACAAGTAGGTCGAGATGTATGCTATGTGCAGTTC 1547
              || || |||||  |||| ||||||  |  |||  |  ||| |||| || |||||
Db       2054 GAAGCTATGTGCTTCCTTATGGACCCTAACCTAGGAAGGAGTGTCTGCTACGTCCAGTTT 2113

Qy       1548 CCTCAGAGGTTTGATGGCATAGATAAGAGTGATCGCTACGCCAATCGTAACGTAGTTTTC 1607
              || |||| || |||||:| || |||||  || |||||  | |||  || ||
Db       2114 CCCCAGAGATTCGATGGCATTGACAGGAATGATCGATATGCCAACAGGAACACCGTGTTT 2173

Qy       1608 TTTGATGTTAACATGAAAGGGTTGGATGGCATTCAAGGACCAGTATACGTAGGAACTGGT 1667
              || ||| |||:| ||| ||  | |||||||| |||||||||| || || ||||||||
Db       2174 TTCGATATTAACTTGAGAGGTCTTGATGGCATCCAAGGACCAGTTTATGTCGGAACTGGC 2233

Qy       1668 TGTGTTTTCAACAGGCAAGCACTTTACGGCTACGGGCCTCCTTCTATGCCCAGCTTACGC 1727
              ||||'|||||| |   | | || || || || ! ||| ||   || ||  |      |
Db       2234 TGTGTTTTCAACCGAACAGCTCTATATGGTTATGAGCCCCCAATTAAGCAGAAGAAGGGT 2293

Qy       1728 AAGAGAAAGGATTCTTCATCCTGCTTCTCATGTTGCT---GCCCCTCAAAGAAGAAGCCT 1784
                  |   ||   |   ||                  ||   |||||||||   |
Db       2294 GGTTTCTTGTCATCACTATGTGGCGGTAGGAAGAAGGCAAGCAAATCAAAGAAGGGCTCG 2353

Qy       1785 GCTCAAGATCCAGCTGAGGTATACAGAGATGCAAAAAAGAGAGGATCTCAATGCTGCCATA 1844
              |  |   |  ||   ||                  |  ||||  ||
Db       2354 GACAAGAAGAAGTCGCAGAAGCATGTGGACAGTTCTGTGCCAGTATTCAACCTTGAAGAT 2413

Qy       1845 TTTAATCTTACAGAGATTGATAATTATGACGAGCATGAAAGGTCAATGCTGATCTCCCAG 1904
              |     ||      |    ||  | |  | |||  | | || || ||
Db       2414 ATAGAGGAGGGAGTTGAAGCGCTGGATTTGACGACGAGAAATCACTTCTTATGTCTCAA 2473

Qy       1905 TTGAGCTTTGAGAAAACTTTTGGCTTATCTTCTGTCTTCATTGAGTCTACACTAATGGAG 1964
              |||||  | |:||| |  |  |  ||  ||| ||  || || || ||||||
Db       2474 ATGAGCCTGGAGAAGAGATTTGGCCAGTCCGCAGCGTTTGTTGCCTCCACTCTGATGGAG 2533

Qy       1965 AATGGAGGAGTACCCGAGTCTGCCAACTCACCACCATTCATCAAGGAAGCGATTCAAGTC 2024
              |||| || || ||  |||| || |    |  |  | || |||||  || || ||
Db       2534 TATGGTGGTGTTCCTCAGTCCGCAACTCCGGAGTCTCTTCTGAAAGAAGCTATCCATGTT 2593

Qy       2025 ATCGGCTGTGGCTATGAAGAGAAGACTGAATGGGGAAAACAGATTGGTTGGATATATGGG 2084
              ||  |||||||||||||| || |||||||||||||||    |||| || ||||| || ||
Db       2594 ATAAGCTGTGGCTATGAGGACAAGACTGAATGGGGAACTGAGATCGGGTGGATCTACGGT 2653

Qy       2085 TCAGTCACTGAGGATATCTTAAGTGGCTTCAAGATGCACTGCCGAGGATGGAGATCAATT 2144
              || || || || || ||  | |  || |||||||||||||    ||||| ||| | || ||
Db       2654 TCTGTGACAGAAGACATTCTCACCGGATTCAAGATGCACGCGCGAGGCTGGCGGTCGATC 2713

Qy       2145 TACTGCATGCCCGTAAGGCCTGCATTCAAAGGATCTGCACCCATCAACCTGTCTGATAGA 2204
              ||||||||||||    || | || |||| || ||||| |||||||| || || |
Db       2714 TACTGCATGCCCAAGCGGCCAGCTTTCAAGGGGTCTGCCCCCATCAATCTTTCGGACCGT 2773

Qy       2205 TTGCACCAGGTCCTCCGATGGGCTCTTGGTTCTGTGGAAATTTTCTTTAGCAGACACTGT 2264
              || :.|||| |||:|  |||||||:|| || ||||| ||  |||| ||| ! |||||
Db       2774 CTGAACCAGGTGCTCCGTTGGGCTCTTGGGTCCGTGGAGATCCTCTTCAGCCGGCACTGC 2833
```

```
2265 CCCCTCTGGTACGGGTTTGGAGGAGGCCGTCTTAAATGGCTCCAAAGGCTTGCGTATATA 2324
2834 CCCCTGTGGTACGGCT---ACGGAGGGCGGCTCAAGTTCCTGGAGAGATTCGCGTACATC 2890

2325 AACACCATTGTGTACCCATTTACATCCCTCCCTCTCATTGCCTATTGCACAATTCCTGCA 2384
2891 AACACCACCATCTACCCGCTCACGTCCATCCCGCTTCTCATCTACTGCATCCTGCCGCC 2950

2385 GTTTGTCTGCTCACCGGAAAATTCATCATACCAACGCTCTCAAACCTGGCAAGCATGCTG 2444
2951 ATCTGTCTGCTCACCGGAAAGTTCATCATTCCAGAGATCAGCAACTTCGCCAGCATCTGG 3010

2445 TTTCTTGGCCTCTTTATCTCCATCATTGTAACTGCGGTGCTTGAGCTAAGATGGAGCGGT 2504
3011 TTCATCTCCCTCTTCATCTCGATCTTCGCCACGGGCATCCTGGAGATGAGGTGGAGCGGG 3070

2505 GTCAGCATTGAAGATTTATGGCGTAATGAACAATTCTGGGTGATCGGAGGTGTTTCAGCC 2564
3071 GTGGGCATCGACGAGTGGTGGAGGAACGAGCAGTTCTGGGTGATCGGGGGCATCTCCGCG 3130

2565 CATCTCTTTGCGGTCTTCCAGGGATTCTTAAAAATGTTGGCTGGCATCGATACGAACTTC 2624
3131 CACCTCTTCGCCGTGTTCCAGGGCCTGCTCAAGGTGCTGGCCGGCATCGACACCAACTTC 3190

2625 ACTGTCACAGCAAAAG---CAGCCGAAGATGCAGAATTTGGGGAGCTATATATGGTCAAG 2681
3191 ACCGTCACCTCCAAGGCCTCGGACGAGGACGGCGACTTCGCGGAGCTGTACATGTTCAAG 3250

2682 TGGACAACACTTTTGATTCCTCCAACCACACTTCTCATTATCAATATGTCGGG---TTGT 2738
3251 TGGACGACGCTCCTGATCCCGCCCACCACCATCCTGATCATCAACCTGGTCGGCGTCGTC 3310

2739 GCTGGATTCTCTGATGCACTCAACAAAGGATATGAAGCATGGGGGCCTCTCTTTGGCAAG 2798
3311 GCCGGCATCTCCTACGCCATCAACAGCGGATACCAGTCGTGGGGCCCGCTCTTCGGCAAG 3370

2799 GTGTTCTTTGCTTTCTGGGTGATTCTTCATCTCTATCCATTCCTTAAAGGTCTAATGGGT 2858
3371 CTCTTCTTCGCCTTCTGGGTCATCGTCCACCTGTACCCGTTCCTCAAGGGCCTCATGGGC 3430

2859 CGCCAAAACCTAACACCAACCATTGTTGTTCTCTGGTCAGTGCTGTTGGCCTCTGTCTTC 2918
3431 AGGCAGAACCGCACCCCGACCATCGTCGTCGTCTGGGCCATCCTGCTGGCGTCCATCTTC 3490

2919 TCTCTCGTTTGGGTCAAGATCAATCCATTC 2948
3491 TCCTTGCTGTGGGTTCGCATCGACCCCTTC 3520
```

```
ESULT 4
Y104730
OCUS       AY104730               3783 bp    mRNA    linear   HTC 16-OCT-2002
EFINITION  Zea mays PCO100501 mRNA sequence.
CCESSION   AY104730
ERSION     AY104730.1  GI:21207808
EYWORDS    HTC.
OURCE      Zea mays
 ORGANISM  Zea mays
           Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
           Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
           clade; Panicoideae; Andropogoneae; Zea.
EFERENCE   1  (bases 1 to 3783)
 AUTHORS   Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S.,
           Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
 TITLE     Maize Mapping Project/DuPont Consensus Sequences for Design of
           Overgo Probes
 JOURNAL   Unpublished (2002)
EFERENCE   2  (bases 1 to 3783)
 AUTHORS   Coe,E.H.
 TITLE     Direct Submission
 JOURNAL   Submitted (25-APR-2002) Maize Mapping Project, University of
           Missouri, Columbia, MO 65211, USA
OMMENT     If you are interested in getting corresponding physical clones,
           these are publicly available from ZmDB and may be found by BLAST
           searching at MSU: maizemap.org, ZmDB, www.zmdb.iastate.edu; TIGR,
           www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the
           maize cDNA sequences is either Virginia Walbot, Stanford or Pat
           Schnable, Iowa State, then clones may be requested from ZmDB:
           www.zmdb.iastate.edu.
FEATURES             Location/Qualifiers
     source          1..3783
                     /organism="Zea mays"
                     /mol_type="mRNA"
                     /db_xref="MaizeDB:635[illegible]"
                     /db_xref="taxon:4577"
                     /clone_lib="Maize Mapping Project/DuPont Consensus
                     Library"
                     /note="This sequence is part of a project of EST
                     assembl[illegible] resulting from the application of public
                     contig[illegible] to some DuPont contigs; this resource was
                     assembl[illegible] by DuPont as part of a collaboration for the
                     [illegible] in conjunction with the Maize
                     Mapp[illegible]"
BASE COUNT       [illegible]
ORIGIN

  Query Match             [illegible]%;  Score [illegible];  DB 11;  Length 3783;
  Best Local Similarity   [illegible]%;  Pred. No. 6.1e-143;
  Matches 1744;  Conservative    0;  Mismatches 832;  Indels   [illegible];  Gaps    8;

Qy        385 ATGAATATGGGAA[illegible]AATTTGGAAGAA[illegible]GTTGAAGAGCTGTAAGGATAAAGAGAACA 444
Db        817 ATG[illegible]TACGGGAG[illegible]TAGCATGGAA[illegible]AGGATTGGAGAGCTGGAAGCAGAAGCAGGAGA 876

Qy        445 AGAAGAAAAAGAGAA[illegible]TCCTAAG[illegible]TGAAA[illegible]GAACCAGCTCAAGTTCCT[illegible]AGAACAGC 504
Db        877 GGATGCACCAGAC[illegible]AAGGAACGATC[illegible]GGC[illegible]GATGATGGTGATGATGCAGATCTACCAC 936

Qy        505 AGATGGAAGAGAAA[illegible]TGTCTGCAGA[illegible]GCTTC[illegible]AGCCGCTTTCAATTGTTTATCCAAT[illegible] 564
Db        937 TAATGGA------[illegible]---TCAAGCTAGAC[illegible]AGCCATTGTCCAGAAAGATCCCGCTTC 981

Qy        565 CACGCAACAAGCTCA[illegible]ACCATACAGAGCAGTGATC[illegible]ATTATGCGACTGGTCATTCTGGGCC 624
Db        982 CTTCAAGCCAAACT[illegible]AACCCCTATAGTAGTGATTA[illegible]AATAATTCGGCTAGTGGTTTTGTGTT 1041

Qy        625 TCTTCTTCCACTT[illegible]AAATAACAAATCC[illegible]TGTAATAGTGCCTTTGGC[illegible]TGTGGCTTACTT 684
Db       1042 TCTTCTTCCACTA[illegible]AGTGATGCATC[illegible]TCTGATGCATTTGCTTTATAGGCTCATAT 1101

Qy        685 CTGTCATATGTGAGATCTGGTTTGCATTTTCTTGGGTGTTGGATCAGTTCCCCAAGTGGA 744
Db       1102 CTGTGATCTGTGAAA[illegible]TTGGTTTGC[illegible]ATGTCTTGGATTCTTGACCAGTTTCCAAAGTGGT 1161

Qy        745 ATCCTGTCAATAGA[illegible]AAACGTATAT[illegible]GAAAG[illegible]TGTCGGCAAGGTATGAAAGAGAGGGTG 804
Db       1162 TTCCTATCGAGAG[illegible]AAACCTATC[illegible]TGAGTTTAAGGTTTGACAAGGAAGGGC 1221

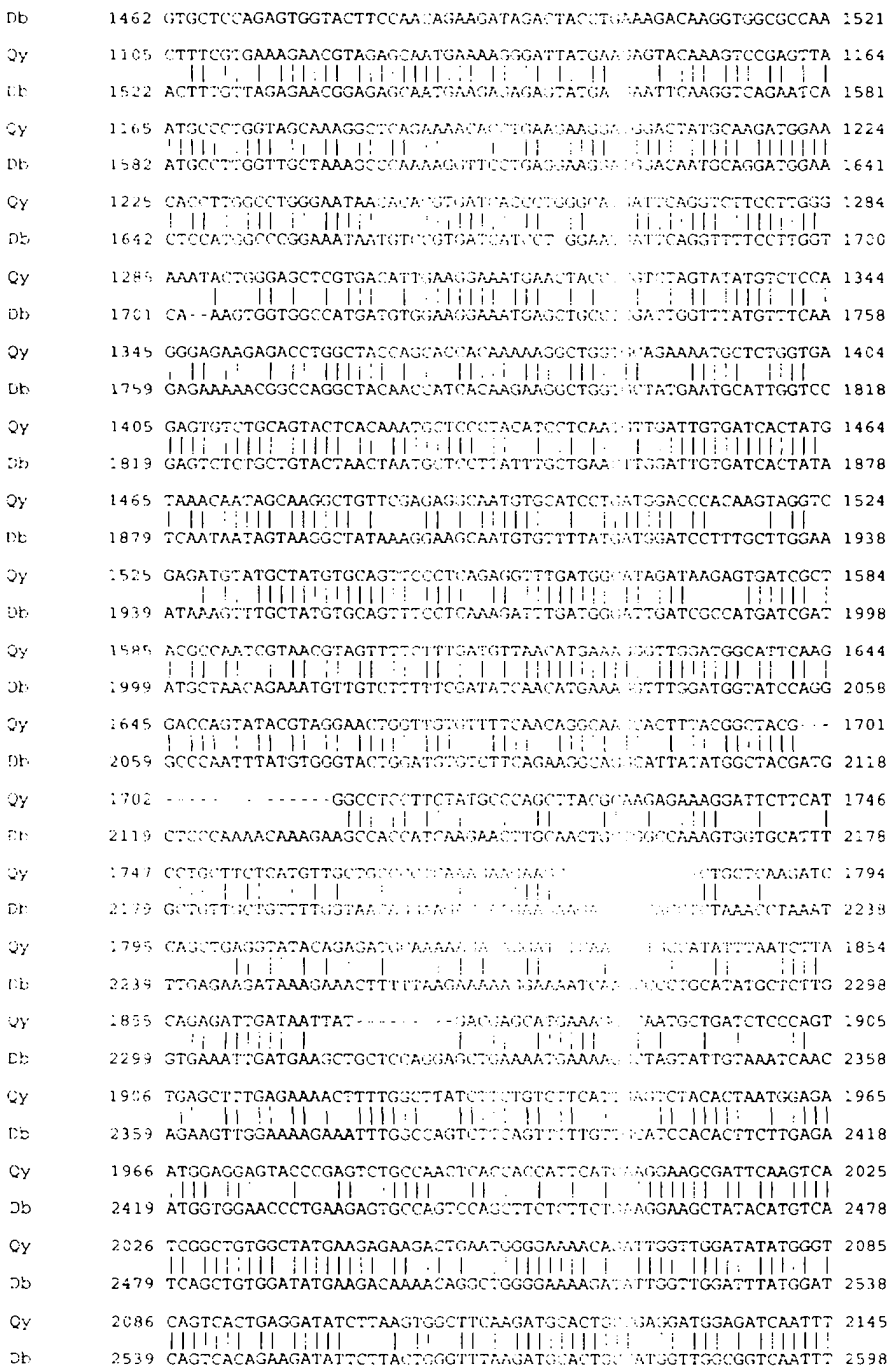
Qy        805 AGCCTTCTCAGCTT[illegible]TGGTGTGGATTTTTT[illegible]TGAGTACTGTTGATCCGCTGAAGGAAC 864
Db       1222 ATCCTTCTCAACT[illegible]CCTGTGGATTT[illegible]AGTACGGTTGATCCCTTGAAGGAAC 1281

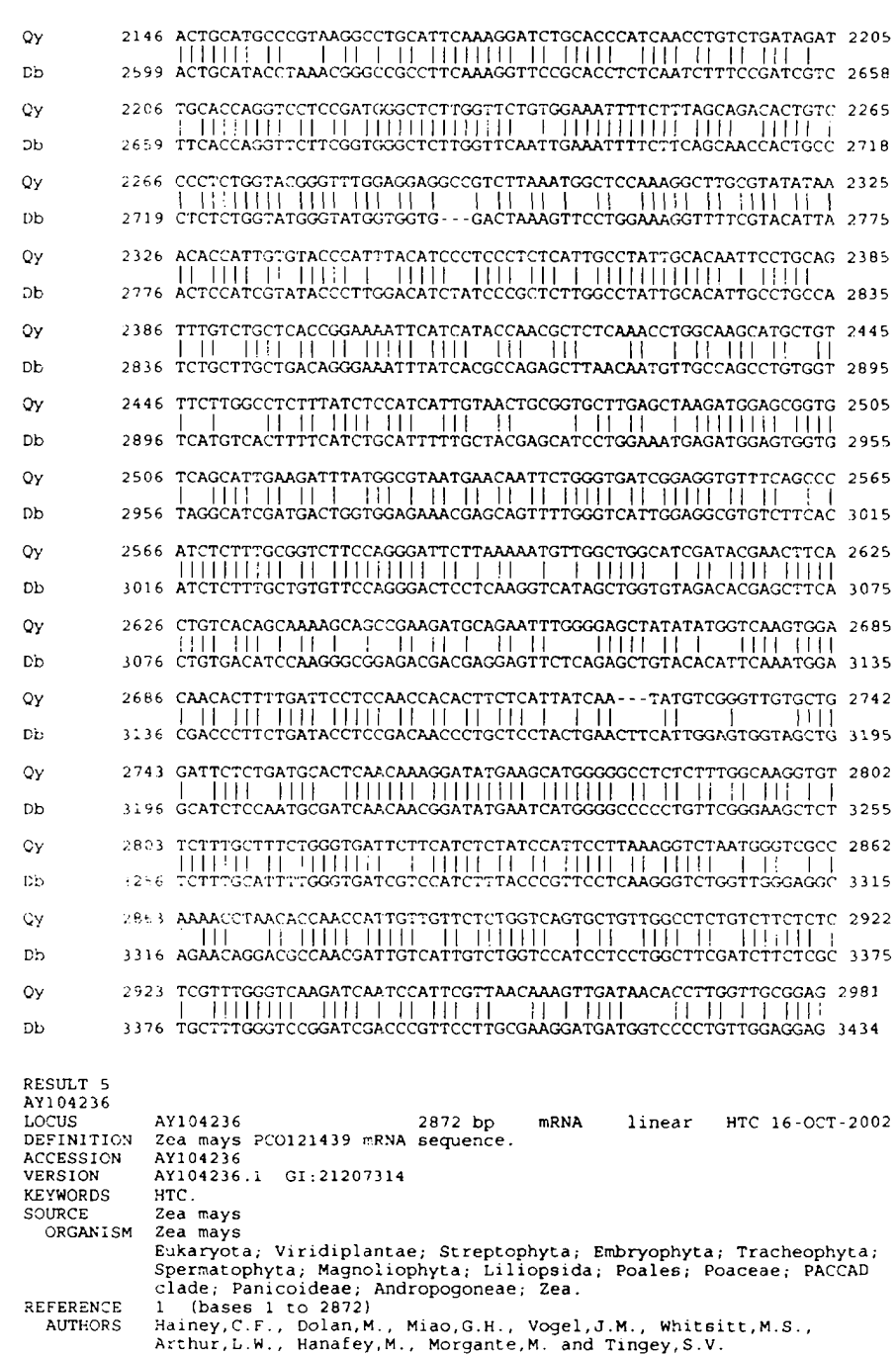
Qy        865 CGCCATTGATCACT[illegible]AATACAGTCCTTTCCATCCTTGCTGTGGACTATCCGTCCATA 924
Db       1282 CTCCATTGGTCAC[illegible]TAATACTGTTCTATCTATCCTTTCGGTGGATTATCCAGTTGATA 1341

Qy        925 AAGTCTCCTGCTAC[illegible]TGTCTGATGATGGTGCAGCTATGCTTTCATTTGAATCTCTTGTAG 984
Db       1342 AGGTTTCATGCTACGTTTCTGATGATGGTTGCTGCATGCTGACATTTGAAGCATTGTCTG 1401

Qy        985 AAACAGCTGAGTTTGCAAGGAAGTGGGTTCCGTTCTGCAAAAAATTCTCAATTGAACCAA 1044
Db       1402 AAACATCTGAATTTGCAAAGAAATGGGTTCCTTTCTGCAAAAGATATAGCCTTGAGCCTC 1461

Qy       1045 GAGCACCGGAGTTTTACTTCTCACAGAAAATTGATTACTTGAAAGACAAGGTTCAACCTT 1104
```





```
RESULT 5
AY104236
LOCUS       AY104236               2872 bp    mRNA    linear   HTC 16-OCT-2002
DEFINITION  Zea mays PCO121439 mRNA sequence.
ACCESSION   AY104236
VERSION     AY104236.1  GI:21207314
KEYWORDS    HTC.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 2872)
  AUTHORS   Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S.,
            Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
```

```
TITLE     Maize Mapping Project/DuPont Consensus Sequences for Design of
          Overgo Probes
JOURNAL   Unpublished (2002)
FERENCE   2  (bases 1 to 2872)
AUTHORS   Coe,E.H.
TITLE     Direct Submission
JOURNAL   Submitted (25-APR-2002) Maize Mapping Project, University of
          Missouri, Columbia, MO 65211, USA
MMENT     If you are interested in getting corresponding physical clones,
          these are publicly available from ZmDB and may be found by BLAST
          searching at MSL, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR,
          www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the
          maize cDNA sequences is either Virginia Walbot, Stanford or Pat
          Schnable, Iowa State, then clones may be requested from ZmDB:
          www.zmdb.iastate.edu.
ATURES              Location/Qualifiers
   source           1. .2872
                    /organism="Zea mays"
                    /mol_type="mRNA"
                    /db_xref="MaizeDB:637064"
                    /db_xref="taxon:4577"
                    /clone_lib="Maize Mapping Project/DuPont Cornsensus
                    Library"
                    /note="this sequence is part of a project of EST
                    assemblies resulting from the application of public
                    contigs to seed DuPont contigs; this resource was
                    assembled by DuPont as part of a collaboration for the
                    overgo addressing of BACs in conjunction with the Maize
                    Mapping Project"
SE COUNT       730 a    600 c    688 g    852 t      2 others
IGIN

Query Match             33.8%;  Score 1093.4;  DB 11;  Length 2872;
Best Local Similarity   68.1%;  Pred. No. 9.9e-142;
Matches 1639;  Conservative    1;  Mismatches  747;  Indels   18;  Gaps    8;

        557 TCCAATTCCACGCAACAAGCTCACACCATACAGAGCAGTGATCATTATGCGACTGGTCAT 616
            ||| ||  | |  ||| |||| |  |  ||  |    |||||  || | || ||  ||||
)        39 TCCGATATCTCCAAACGAGCTTAACCTTTATCGGATCGTGATTGTTCTCCGGCTTATCAT  98

/       617 TCTGGGCCTCTTCTTCCACTTCAGAATAACAAATCCTGTCGATAGTGCCTTTGGCCTGTG 676
            ||  |  ||||||| || |   | |||||  |||| || ||   ||| | |||  ||||
)        99 CCTATGTTTCTTCTTTCAATATCGTATAACTCATCCAGTGGAAGATGCTTATGGGTTGTG 158

/       677 GCTTACTTCTGTCATATGTGAGATCTGGTTTGCATTTTCTTGGGTGTTGGATCAGTTCCC 736
            ||||  |||||  || |||||  | |||||||| || ||||||  |  | |||||.||||
)       159 GCTTGTATCTGTTATTTGTGAAGTTTGGTTTGCCTTGTCTTGGCTTCTAGATCAGTTCCC 218

/       737 CAAGTGGAATCCTGTCAATAGAGAAACGTATATCGAAAGGCTGTCGGCAAGGTATGAAAG 796
            ||||||  ||||| ||||  |  ||||| ||  |||| || ||  |    || ||||| ||
)       219 AAAGTGGTATCCTATCAACCGTGAAACTTACCTCGATAGACTTGCATTGAGATATGATAG 278

/       797 AGAGGGTGAGCCTTCTCAGCT-TGCTGGTGTGGATTTTTTCGTGAGTACTGTTGATCCGC 855
            ||||||||||||  || |||  |  |||     | ||| | || || ||||| || |||| |
)       279 GGAGGGTGAGCCATCCCAGTTGGGCTCCAATCGATGTCTTTGTTAGTACAGTGGATCCAC 338

/       856 TGAAGGAACCGCCATTGATCACTGCCAATACAGTCCTTTCCATCCTTGCTGTGGACTATC 915
            | |||||||| ||  | || |||| ||| || ||||| |||| ||||||||||| || | |
)       339 TTAAGGAACCTCCTCTAATTACTGGCAACACTGTCCTGTCCATTCTTGCTGTGGATTACC 398

y       916 CCGTCGATAAAGTCTCCTGCTACGTGTCTGATGATGGTGCAGCTATGCTTTCATTTGAAT 975
            | || || ||||| || || || || |||||||| ||| |||||||| |  | |||||
b       399 CTGTTGACAAAGTATCATGTTATGTTTCTGATGACGGTTCAGCTATGTTGACTTTTGAAG 458

y       976 CTCTTGTAGAAACAGCTGAGTTTGCAAGGAAGTGGGTTCCGTTCTGCAAAAAATTCTCAA 1035
            | ||    ||||| || ||||||||||||||| |||||||| || ||||| |||  |   |
b       459 CGCTATCTGAAACCGCAGAGTTTGCAAGGAAATGGGTTCCCTTTTGCAAGAAACACAATA 518

y      1036 TTGAACCAAGAGCACCGGAGTTTTACTTCTCACAGAAAATTGATTACTTGAAAGACAAGG 1095
            |||||| || || || |||||||||||  | |  || || |||||| | || |||||
b       519 TTGAACCTAGGGCTCCAGAGTTTTACTTTGCTCGAAAGATAGATTACCTAAAGGACAAAA 578

y      1096 TTCAACCTTCTTTCGTGAAAGAACGTAGAGCAATGAAAAGGGATTATGAAGAGTACAAAG 1155

Db      579 TACAACCTTCTTTC TGAAAGAAAGAAGGCTATGAAGAGGGAGTGTGAAGAGTTCAAAG 638

Qy     1156 TCCGAGTTAATA TGTAGCAAAGGCTCAGAAAACACCTGAAGAAGGATGGACTATGC 1215
            | ||  ||| ||  | ||  | |  | ||| | ||  | ||| | || | |||| ||
Db      639 TACGGATCGATG TTGTTGCAAAAGGCAAAAAATACCTGAGGAGGGCTGGACCATGG 698

Qy     1216 AAGATGGAACAC GCCTGGGAACAACACAGTGATCACCCTGGGCATGATTCAGGTC 1275
            |  ||  |  ||  ||| |||  |||    | |||| ||  || |||| |  ||
Db      699 CTGATGGCACT CCTGGAATAACCCTAGAGATCATCC-MGGAATGATCCAAGTA 757

Qy     1276 TTCCTTGGGAAATA GGACTCGGACATTGAAGGAAATGAACTACCTCGTCTAGTAT 1335
            ||| | ||      |  | |  |  ||| | |. |||||  | || ' | |  | |
Db      758 TTCTTGGGCCATA ATGGTCT TGACACGATGGGAATGAGTTGCCACGGCTTGTTT 815

Qy     1336 ATGTCTCCAGGGA AATAGACCTGGTTACCAGCACCACAAAAAGGCTGGTGCAGAAAATG 1395
            |  | |  |     |  |  |  ||  ||| ||||| .|||'|.|||    ||  |
Db      816 ATGTTTCTCGT AAA AAAGGCAGTTTTACTCACTCACAAGAAGGCTGGTGCCATGAATG 875

Qy     1396 CTCTGGTGAGA      ACTACATCCTCAATGTTGATTGTG 1455
            ||  | |  |      |  ||  | ||  | ||  | ||:|||
Db      876 CTTTGATTCGC     TTATGTTCTTAACGTGGATTGTG 935

Qy     1456 ATCACTATGTAAA AATAGCAAGCCGTTCGAGAGGCAATGTGCATCCTGATGGACCCAC 1515
            |||||| | ||   ||  | |  ||  ||| .|||| '| |||.||| ||
Db      936 ATCACTACTTCAATAGCAGCAAAGCTCTTAGAGAGGCTATGTGTTTCATGATGGATCCAG 995

Qy     1516 AAGTAGGTCGAGAT TATGCTATGTTGCAGTTCCCTCAGAGGTTTGATGGCATAGATAAGA 1575
            . |||  |  |    ||||||||  .||| :  || .| ||||||| || | |  |
Db      996 CACTAGGAAGGAAAAGTTGCTATGTTCAGTTTCCACAAAGATTTGATGGTATAGACTTGC 1055

Qy     1576 GTGATCGCTACGGCAATCGTAACGTAGTTTTCTTTGATGTTAACATGAAAGGGTTGGATG 1635
            |||||| ||  | | | |.| | |  |||| ||| ||.| |||| || |  ||||
Db     1056 ATGATCGATATGCAAACGGAACATTGTCTTCTTTGATATTAATATGAAGGGTCTAGATG 1115

Qy     1636 GCATTCAAGGACCA TATACGTAGGAACTGGTTGTGTTTTCAACAGGCAAGCACTTTACG 1695
            ||| |||'.|||.  | || ||  ||| '  ||  '|||.| ||||' ||  .  | ||
Db     1116 GCATTCAAGGACCT TTTATGTGGGAACAGGATGCTGTTTCAATAGGCAGGCTTGTATG 1175

Qy     1696 GCTACGGGGCCTCCTT TATGCGCAGCTTACGCAAGAGAAAGGATTCTTCATCCTGCTTCT 1755
            '||. |  ||'  |                 ||  ||  |          |  ||  ||
Db     1176 GCTATGATCCTGTAT GACAGAAGCTGATTTGAGCCTAACATTATCATTAAAAGTTGCT 1235

Qy     1756 CATGTTGCTGCCCTT AAAGAAGAAGCCTGCTCAAGATCC- --AGCTGAGGTATACAGA 1811
            |  |  |    '||  |||  |  |||  |  | | | | ''| ||'
Db     1236 GTGGCGGAAGAAAAAATGAAGGACAAGAGCTATATTGGATTCCAAAAACCGTGATATGAAGA 1295

Qy     1812 GATGCAAAAAGAGAAGGATCTCAATGCTGCTATATTTAATCTTACAGAGATTGATAATTAT 1871
            ||  || ||    ||  ||  .||  || ;  |||:          |  |
Db     1296 GA ACAGAATCTT GCTCCCATCTTCAACATGGAAGATATAGAAGAGGGATTTGAAGGT 1354

Qy     1872 GACGAGCATGAAAAGTTCAATGCTGATTTCCCAGTTGAGCTTTGAGAAAACTTTTGGCTTA 1931
            |||;| |'|||  '|| | |'||  | ||' |  |'||| ||||||    ||||||
Db     1355 TACGAGGATGAAAA TCACTGCTTATGTCTCAGAAGAGCTTGGAGAAACGCTTTGGCCAG 1414

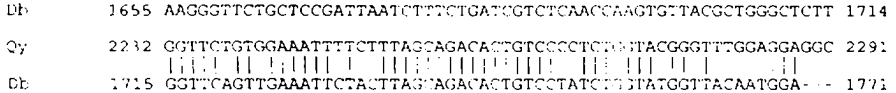
Qy     1932 TCTTCTGTCTTCAT AGTCTACACTAATGGAGAATGGAGGAGTACCCGAGTCTGCCAAC 1991
            ||| |  | ||  ||  || |  | '||    | || ||  ||||   |  | .||
Db     1415 TCTCCAATTTTCAT GCATCTACTTTATGACTCAAGGTGGCATACCCCCTTCAACAAAC 1474

Qy     1992 TCACCACCATTCAT AAGGAAGCGATTCAAGTCATCGGCTGTGGCTATGAAGAGAAGACT 2051
            ||   ,  |  ;  |||| |  ||  |  ||;|  | ;|||| ;|||  ,| |  |
Db     1475 CCAGGTTCCCTGCT AAAGGAAGCTATACATGTCATTAGTTGTGGATATGAGGATAAAACA 1534

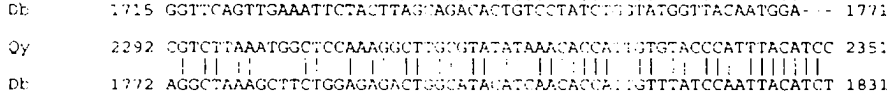
Qy     2052 GAATGGGGAAAACA ATTGGTTGGATATATGGGTCAGTCACTGAGGATATGTTAAGTGGC 2111
            ||.|||  || |||  || |||||| |||  || |  |||| |||| ||| ||
Db     1535 GAATGGGGGAAAGA TCGGATGGATATATGGCTCTGTTACTGAAGATATTTTAACTGGT 1594

Qy     2112 TTCAAGATGCACTGCCGAGGATGGAGATCAATTTACTGCATGCCCGTAAGGCCTGCATTC 2171
            |||||||||||      |  | |'|  || ||  ||||||||||    |  |||'|   |.|
Db     1595 TTCAAGATGCATGCAAGAGGTTGGATATCCATCTACTGCATGCCACTTCGGCCTTGCTTC 1654

Qy     2172 AAAGGATCTGCACCCATCAACCTGTCTGATAGATTGCACCAGGTCCTCCGATGGGCTCTT 2231
            || || ||||| |  || || ||  ' |||| .  |  ;|||| ||  |  | |||'||||
```



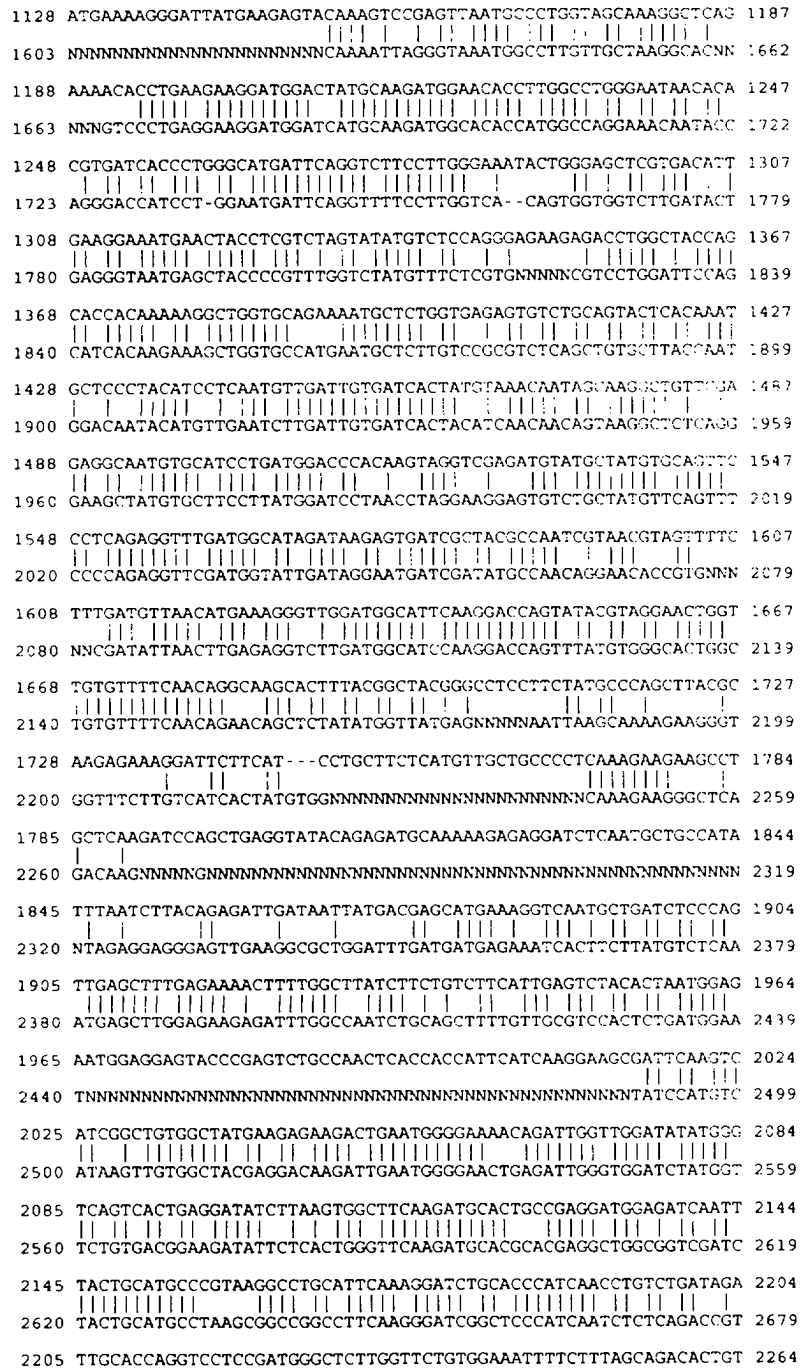
RESULT 6
AY112236
LOCUS AY112236 3728 bp mRNA linear HTC 17-OCT 2002
DEFINITION Zea mays CL1160_1 mRNA sequence.
ACCESSION AY112236
VERSION AY112236.1 GI:21216826
KEYWORDS HTC.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; [illegible]phyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida, [illegible]ales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea
REFERENCE 1 (bases 1 to 3728)
AUTHORS Hainey,C.F., Dolan,M., Miao,G.H., Vaniel,J.S., Whitsitt,M.S., Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
TITLE Maize Mapping Project/DuPont Consensus Sequences for Design of Overgo Probes
JOURNAL Unpublished (2002)

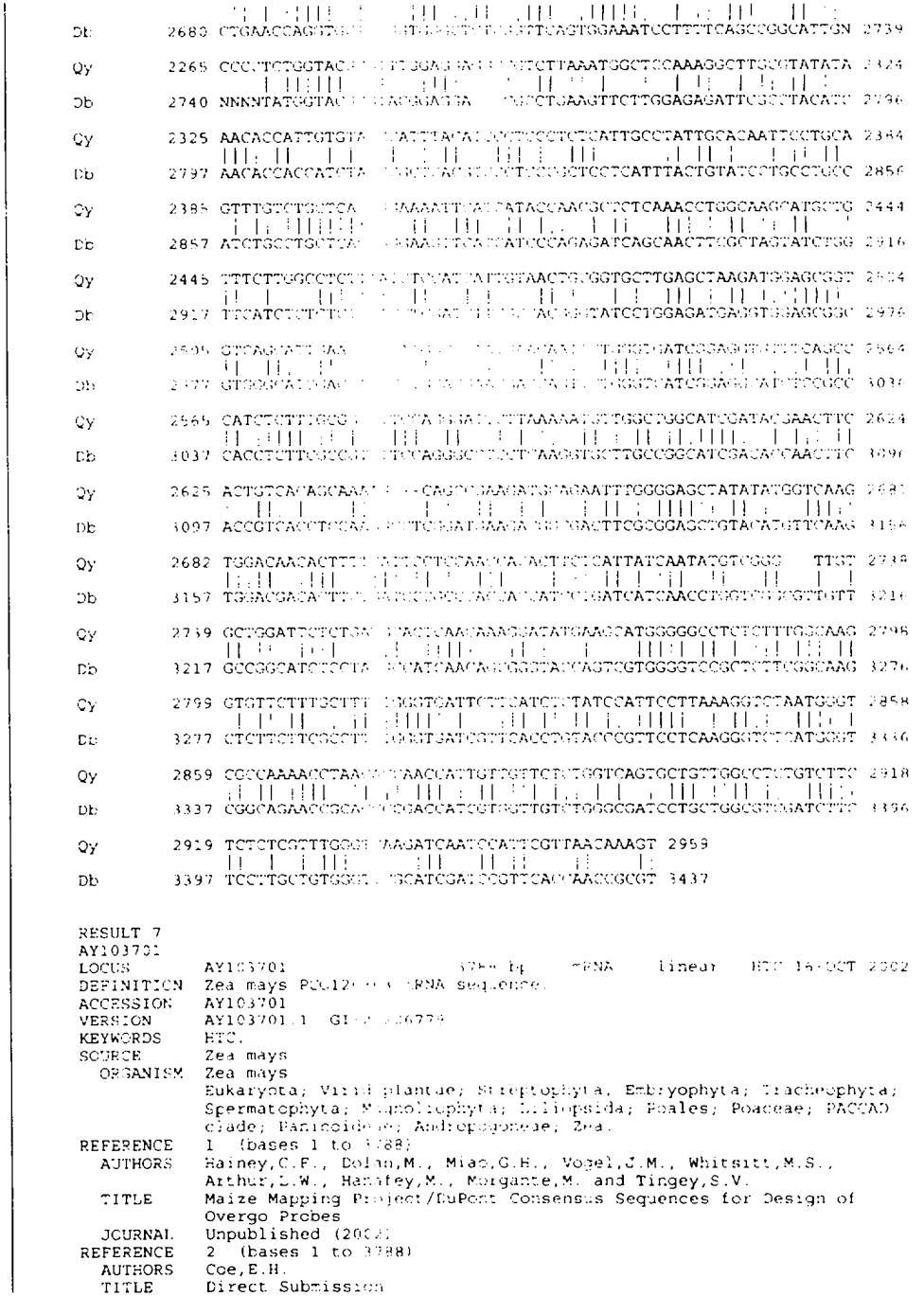


REFERENCE 2 (bases 1 to 3728)
AUTHORS Coe,E.H.
TITLE Direct Submission
JOURNAL Submitted (25-APR-2002) Maize Mapping Project, University of Missouri, Columbia, MO 65211, USA
COMMENT If you are interested in getting corresponding physical clones, these are publicly available from ZmDB and may be found by BLAST searching at MSL, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR, www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the maize cDNA sequences is either Virginia Walbot, Stanford or Pat Schnable, Iowa State, then clones may be requested from ZmDB: www.zmdb.iastate.edu.
FEATURES Location/Qualifiers
source 1..3728
/organism="Zea mays"
/mol_type="mRNA"
/db_xref="MaizeDB:630049"
/db_xref="taxon:4577"
/clone_lib="Maize Mapping Project/DuPont Cornsensus Library"
/note="this sequence is part of a project of EST assemblies resulting from the application of public contigs to seed DuPont contigs; this resource was assembled by DuPont as part of a collaboration for the overgo addressing of BACs in conjunction with the Maize Mapping Project"
BASE COUNT 782 a 845 c 925 g 878 t 298 others
ORIGIN

Query Match 33.6%; Score 1086.8; DB 11; Length 3728;
Best Local Similarity 64.2%; Pred. No. 7.2e-141;
Matches 1568; Conservative 0; Mismatches 858; Indels 15; Gaps 6;







```
RESULT 7
AY103701
LOCUS       AY103701                5784 bp    mRNA    linear   PLN 16-OCT 2002
DEFINITION  Zea mays PCO120[illegible] mRNA sequence.
ACCESSION   AY103701
VERSION     AY103701.1  GI:[illegible]6779
KEYWORDS    HTC.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 5784)
  AUTHORS   Rainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S.,
            Arthur,L.W., Hamaker,M., Morgante,M. and Tingey,S.V.
  TITLE     Maize Mapping Project/DuPont Consensus Sequences for Design of
            Overgo Probes
  JOURNAL   Unpublished (2002)
REFERENCE   2  (bases 1 to 5784)
  AUTHORS   Coe,E.H.
  TITLE     Direct Submission
```

```
  JOURNAL   Submitted (25-APR-2002) Maize Mapping Project, University of
            Missouri, Columbia, MO 65211, USA
COMMENT     If you are interested in getting corresponding physical clones,
            these are publicly available from ZmDB and may be found by BLAST
            searching at MSL, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR,
            www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the
            maize cDNA sequences is either Virginia Walbot, Stanford or Pat
            Schnable, Iowa State, then clones may be requested from ZmDB:
            www.zmdb.iastate.edu.
FEATURES             Location/Qualifiers
     source          1..3788
                     /organism="Zea mays"
                     /mol_type="mRNA"
                     /db_xref="MaizeDB:637005"
                     /db_xref="taxon:4577"
                     /clone_lib="Maize Mapping Project/DuPont Cornsensus
                     Library"
                     /note="this sequence is part of a project of EST
                     assemblies resulting from the application of public
                     contigs to seed DuPont contigs; this resource was
                     assembled by DuPont as part of a collaboration for the
                     overgo addressing of BACs in conjunction with the Maize
                     Mapping Project"
BASE COUNT      851 a    954 c   1114 g    866 t      3 others
ORIGIN

  Query Match             33.1%;  Score 1071.2;  DB 11;  Length 3788;
  Best Local Similarity   66.4%;  Pred. No. 1e-138;
  Matches 1660;  Conservative    0;  Mismatches  796;  Indels   43;  Gaps    7;

Qy        528 GAGGCTTCGGAGCCGCTTTCAATTGTTTATCCAATTCCACG AACAAGCTCACACCATAC 587
              || ||||  | ||||  | || |          || ||  ||   . |    | |  || ||
Db        979 GAAGCTAGGCAGCCATTGTCCAGAAAAGTCGCTATATCATCAAGCCGAATTAATCCCTAC 1037

Qy        588 AGAGCAGTGATCATTATGCGACT GTCATTCTGGGCCTCTT TTCCACTTCAGAATAACA 647
              ||     | ||| |||| |  |||| || |||| ||||  ||| | | | |  |.
Db       1038 AGGATGATTATCGTTATCCGGTTGGTGGTTTTGGGTTCTT TCCACTACCGAGTGATG 1097

Qy        648 AATCCTGTCGATAGTGCCTTTGGCCTGTGGCTTACTTCTT TATATGTGAGATCTGGTTT 707
              ||.|  |   |  ||| |||  .|||.| |  ||||, || |,|'|| |'||||||
Db       1098 CATCCGGCGAAAGATGCATTTGCATTGTGGCTCATATCTG AATCTGTGAAATCTGGTTT 1157

Qy        708 GCATTTTCTTGGGTGTTGGATCAGTTCCCCAAGTGGAATC TGTCAATAGAGAAACGTAT 767
              ||  | || ||| |  | |||'|||'|| ||'|| ||| || |, '|||| || ||
Db       1158 GCGATGTCCTGGATTCTTGATCAGTTCCCAAAGTGGCTT CAATCGAGAGAGAGACTTAC 1217

Qy        768 ATCGAAAGGCTGTCGGCAAGATATAAAGAAAT TTCA    TCAGCTTGCTGGTGTG 827
              | ||  | ||| | ||'  |'  | |  |  |||  |||.
Db       1218 CTGGACCGTTTGTCACTAA TTTA AA AAA TTAA    TCAG TTGCTCCAATC 1277

Qy        828 GATTTTTTCGTGAGTACTGTTGATCCGCTGAAAAAAGTCT  GGATCACTGCCAATACA 887
              | ||  || | ||| |  |||  |  |  || |||  ||  | | |  ||
Db       1278 GACTTCTTTGTCAGTACGGTTGATCCCACAAAGGAACCT TT TGGTCACAGCGAACACT 1337

Qy        888 GTCCTTTCCATCCTTGCTGTGGACTATCCCGTCGATAAAGT TCCTGCTACGTGTCTGAT 947
              ||.|||'|.|||| |||.|| || || || '| ||  |.| '|||' || ||||||
Db       1338 GTCCTTTCCATCCTTTCTGTGGATTATCCGGTTGAGAAGTT TCCTGCTATGTTTCTGAT 1397

Qy        948 GATGGTGCAGCTATGCTTTCACTTGAATCTCTTGTAGAAAC AGTTGAGTTTGCAAGGAAG 1007
              |,|||'| || |||||| | .|||| ' |  | || ||. .||| '|||'| |||
Db       1398 GATGGTGCTGCAATGCTTACGTTTGAAGCATTGTCTGAAAC AGTTGAATTTGCAAAGAAA 1457

Qy       1008 TGGGTTCCGTTCTGCAAAAAATTCTCAATTGAAGCAAGA AAAGGAGTTTACTTCTCA 1067
              |||.|.|| ||| |||||| ||    |. || ||  |  || ||| |||||  |
Db       1458 TGGGTTCCTTTCAGCAAAAAGTT AATATCGAGCCTCGTGCT CTGAGTGGTACTTCCAA 1517

Qy       1068 CAGAAAATTGATTACTTGAAAGACAAGGTTCAAGCTTCTTT TGAAAGAACGTAGAGCA 1127
              |||'. || || ||| |.|||, |||.|'     '||| |. |. |  ||  | || ||
Db       1518 CAGAAGATAGACTACCTGAAAGACAAGGTTGCTGCTTCATT TGTTAGGGAGAGGAGGGCG 1577

Qy       1128 ATGAAAAGGGATTATGAAGAGTACAAAGTCCGAGTTAATGCT TGGTAGCAAAGGCTCAG 1187
              |.|.| || |. || || || |  || ||  |  : ||||.| |||  .||| .| |:
Db       1578 ATGAAGAGAGAATACGAGGAATT AAGGTAAGGATCAAT TGTGGTTGCAAAAGCCCAA 1637

Qy       1188 AAAACACCTGAAGAAGGATGGACTATGCAAGATGGAACACCTTGGCCTGGGAATAACACA 1247
              ||    ||||| |||||| |||| ||||||||||||  || ||||||| || |||  |
Db       1638 AAGGTTCCTGAGGAAGGATGGACAATGCAAGATGGAAGCCCCTGGCCTGGAAACAACGTA 1697

Qy       1248 CGTGATCACCCTGGGCATGATTCAGGTCTTCCTTGGGAAATACTGGGAGCTCGTGACATT 1307
              || |||| ||| || || |||||||| ||||||||  |  |    || | ||||||  |
Db       1698 CGCGATCATCCT-GGAATGATTCAGGTATTCCTTGGCCA--AAGTGGCGGTCGTGATGTG 1754

Qy       1308 GAAGGAAATGAACTACCTCGTCTAGTATATGTCTCCAGGGAGAAGAGACCTGGCTACCAG 1367
              ||||||||||  | ||||| || || ||||||| || || ||||| || || ||  |
Db       1755 GAAGGAAATGAGTTGCCTCGCCTGGTTTATGTCTCGAGAGAAAAGAGGCCAGGTTATAAC 1814

Qy       1368 CACCACAAAAAGGCTGGTGCAGAAAATGCTCTGGTGAGAGTGTCTGCAGTACTCACAAAT 1427
              || ||||| ||||||||||||    ||||| ||||| | || |||||| ||  |  |||||
Db       1815 CATCACAAGAAGGCTGGTGCCATGAATGCACTGGTCCGTGTCTCTGCTGTCTTATCAAAT 1874

Qy       1428 GCTCCCTACATCCTCAATGTTGATTGTGATCACTATGTAAACAATAGCAAGGCTGTTCGA 1487
              ||| | ||| |  | ||  | || |||||||||||  | ||||||||||||||  |  |
Db       1875 GCTGCATACCTATTGAACTTGGACTGTGATCACTACATCAACAATAGCAAGGCCATAAAA 1934

Qy       1488 GAGGCAATGTGCATCCTGATGGACCCACAAGTAGGTCGAGATGTATGCTATGTGCAGTTC 1547
              ||||| ||||| || ||||||| ||      || ||      | || |||||||| ||||||
Db       1935 GAGGCTATGTGTTTCATGATGGATCCTTTGGTGGGGAAGAAAGTGTGCTATGTACAGTTC 1994

Qy       1548 CCTCAGAGGTTTGATGGCATAGATAAGAGTGATCGCTACGCCAATCGTAACGTAGTTTTC 1607
              |||||||||||||||| || || || | |||||| ||||| ||  | ||||| || ||
Db       1995 CCTCAGAGGTTTGATGGTATTGACAAAAATGATCGATACGCTAACAGGAACGTTGTCTTT 2054

Qy       1608 TTTGATGTTAACATGAAAGGGTTGGATGGCATTCAAGGACCAGTATACGTAGGAACTGGT 1667
              |||||  | |||||||||||| |||| || |||||||||||  | || || || |||||
Db       2055 TTTGACATCAACATGAAAGGTTTGGACGGTATTCAAGGACCCATTTATGTGGGTACTGGA 2114

Qy       1668 TGTGTTTTCAACAGGCAAGCACTTTACGGCTACG----------------GGCCTCCT 1709
              ||||||||||   |||| |||| || || || |                ||| ||
Db       2115 TGTGTTTTCAGACGGCAGGCACTGTATGGTTATGATGCTCCTAAAACGAAGAAGCCACCA 2174

Qy       1710 TCTATGCCCAGCTTACGCAAGAGAAAGGATTCTTCATCCTGCTTCTCATGTTGCTGCCCC 1769
              || |  |  ||   ||  |  |||  |   || |||| ||  |  | |  |
Db       2175 TCAAGAACTTGCAACTGCTGGCCCAAGTGGTGCCTCTCTTGCTGCTGCAGCAGGAACAAG 2234

Qy       1770 TCAAAGAAGAAGCCTGCTCAAGATCCAGCTGAG---------------GTATACAGAGAT 1814
              || |||||| || |  ||  |||  |  |||                | | ||  |
Db       2235 AATAAAAAGAAGACTACAAAACCAAAGACGGAGAAGAAGAAAAGATTATTTTTCAAGAAA 2294

Qy       1815 GCAAAAAGAGAGGATCTCAATGCTGCCATATTTAATCTTACAGAGATTGATAATTATGAC 1874
              |,| |||        ||       |||  |  | |  ||  ||  || |      ||
Db       2295 GCAGAAAACCCATCTCCTGCATATGCTTTGGGTGAAATTGATGAAGGTGCTCCAGGTGCT 2354

Qy       1875 GAGCATGAAAGGTCAATGCTGATCTCCCAGTTGAGCTTTGAGAAAACTTTTGGCTTATCT 1934
              :'  || | | |  |  ||   ||  | ||||| |  |||||    |||
Db       2355 GATATCGAGAAGGCCGGAATCGTAAATCAACAGAAACTAGAGAAGAAATTTGGGCAGTCT 2414

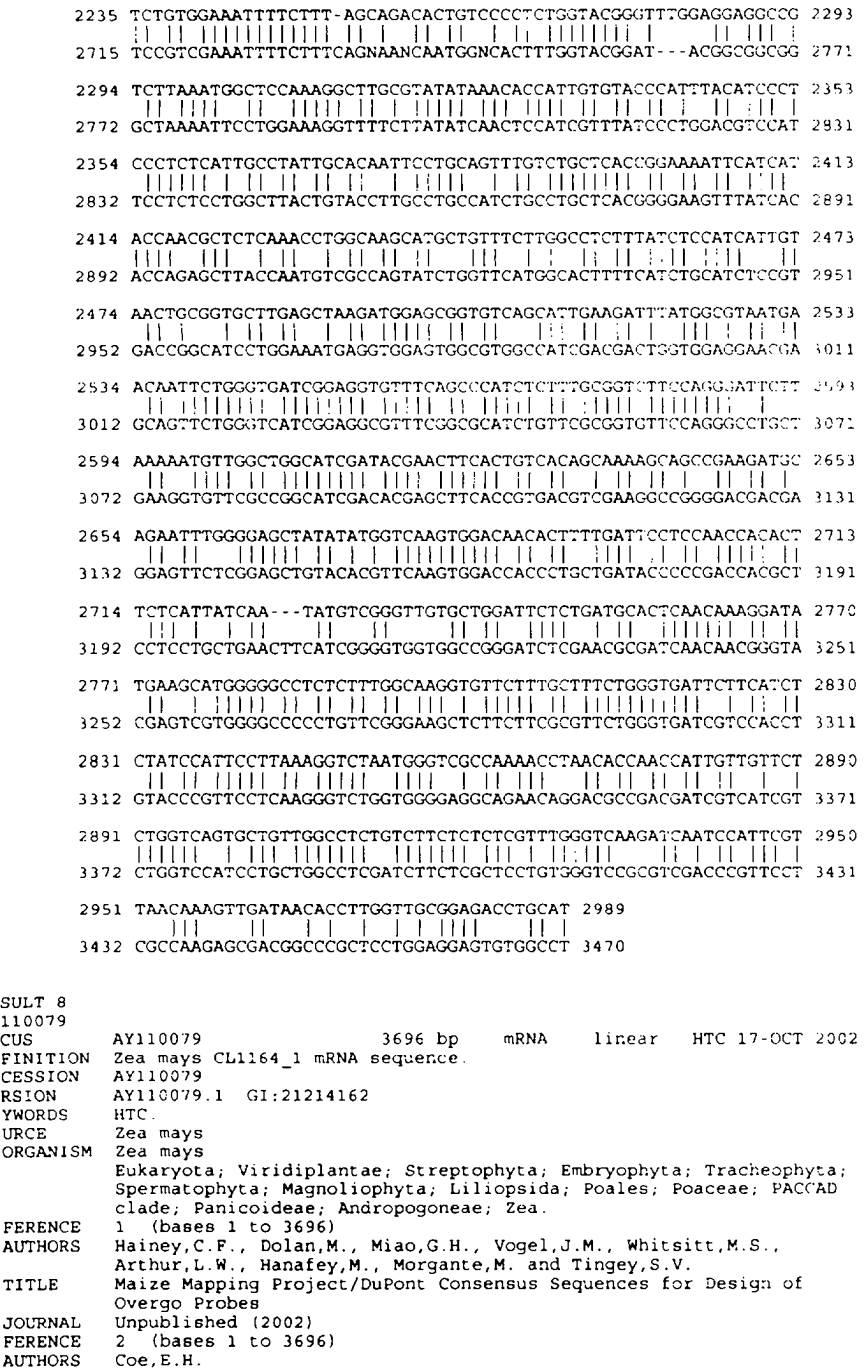
Qy       1935 TCTGTCTTCATTGAGTCTACACTAATGGAGAATGGAGGAGTACCCGAGTCTGCCAACTCA 1994
              ||||| ||  | |  || |||||  | ||||| |||||     |   ||  || |   ||
Db       2415 TCTGTTTTTGTCGCATCAACACTTCTTGAGAACGGAGGGACCCTGAAGAGCGCAAGTCCA 2474

Qy       1995 CCACCATTCATCAAGGAAGCGATTCAAGTCATCGGCTGTGGCTATGAAGAGAAGACTGAA 2054
              |  |  |  | |||||||| || || || ||| |||| ||||| ||||| ||||| ||
Db       2475 GCTTCTCTTCTGAAGGAAGCTATACATGTTATCAGCTGCGGCTACGAAGACAAGACCGAC 2534

Qy       2055 TGGGGAAAACAGATTGGTTGGATATATGGGTCAGTCACTGAGGATATCTTAAGTGGCTTC 2114
              |||||||| |||||| ||||| || || || |||| |||||||||| | ||| ||
Db       2535 TGGGGAAAAGAGATTGGCTGGATTTACGGATCGATCACAGAGGATATCTTGACTGGATTT 2594

Qy       2115 AAGATGCACTGCCGAGGATGGAGATCAATTTACTGCATGCCCGTAAGGCCTGCATTCAAA 2174
              |:|||||||||||  || ||| | || |||||||||| ||     ||||||||||||
Db       2595 AAGATGCACTGCCATGGCTGGCGGTCTATTTACTGCATCCCGAAGCGGCCTGCATTCAAA 2654

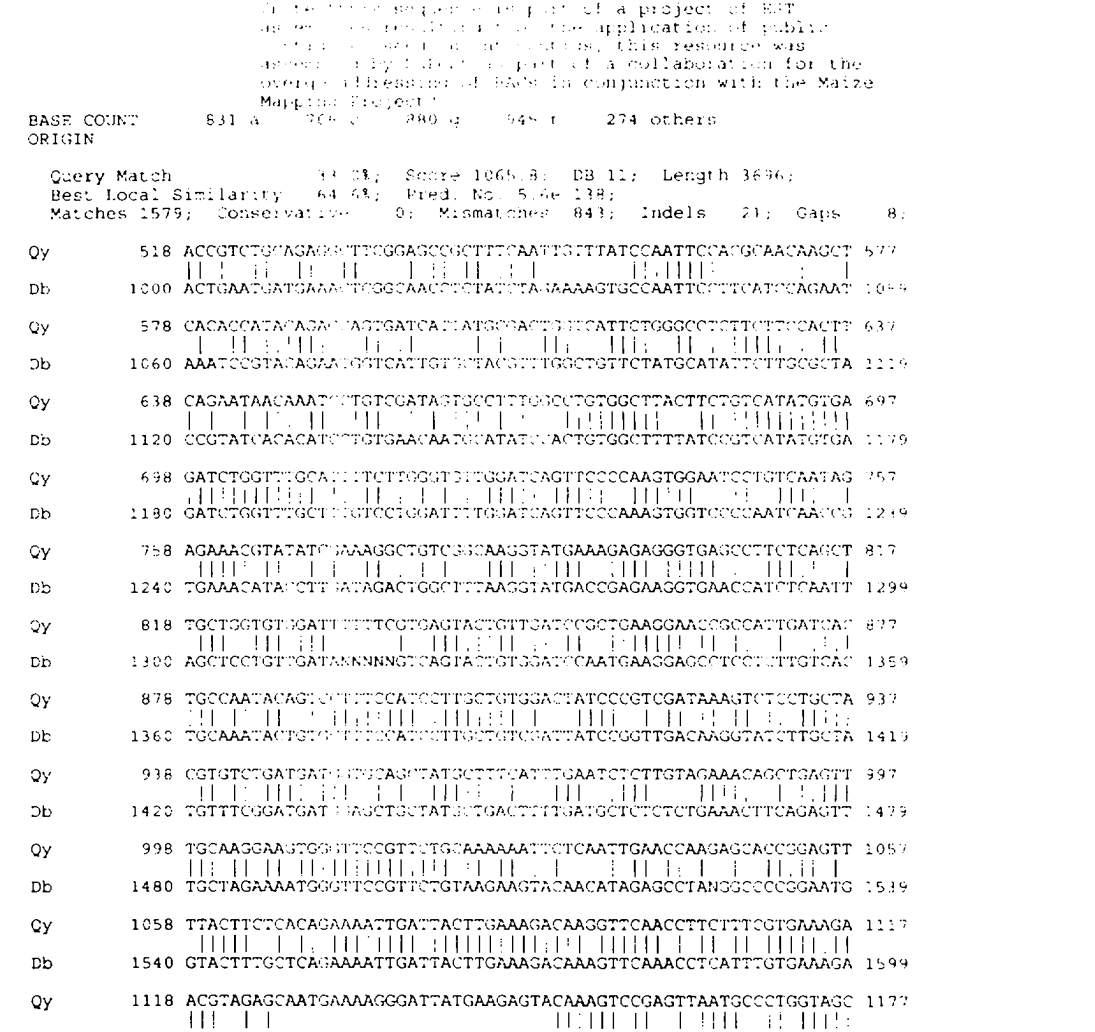
Qy       2175 GGATCTGCACCCATCAACCTGTCTGATAGATTGCACCAGGTCCTCCGATGGGCTCTTGGT 2234
              :  |||:| ||  | |||:| || ||  |  | |||||||||| || |||:| |||||
Db       2655 GGTTCTGCGCCTCTGAACCTTTCCGACCGTCTTCACCAGGTCCTTCGCTGGGCCCTTGGG 2714
```

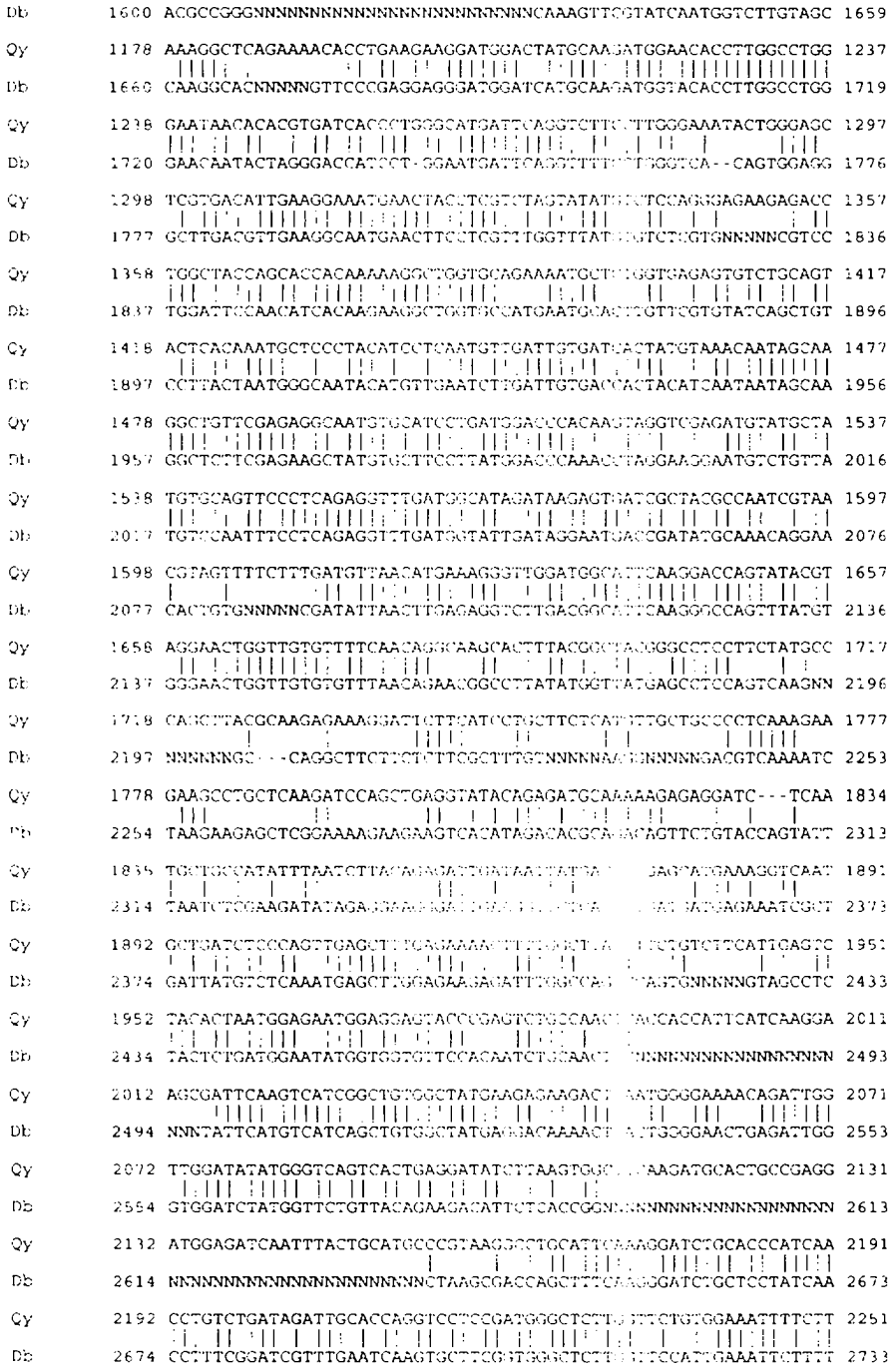


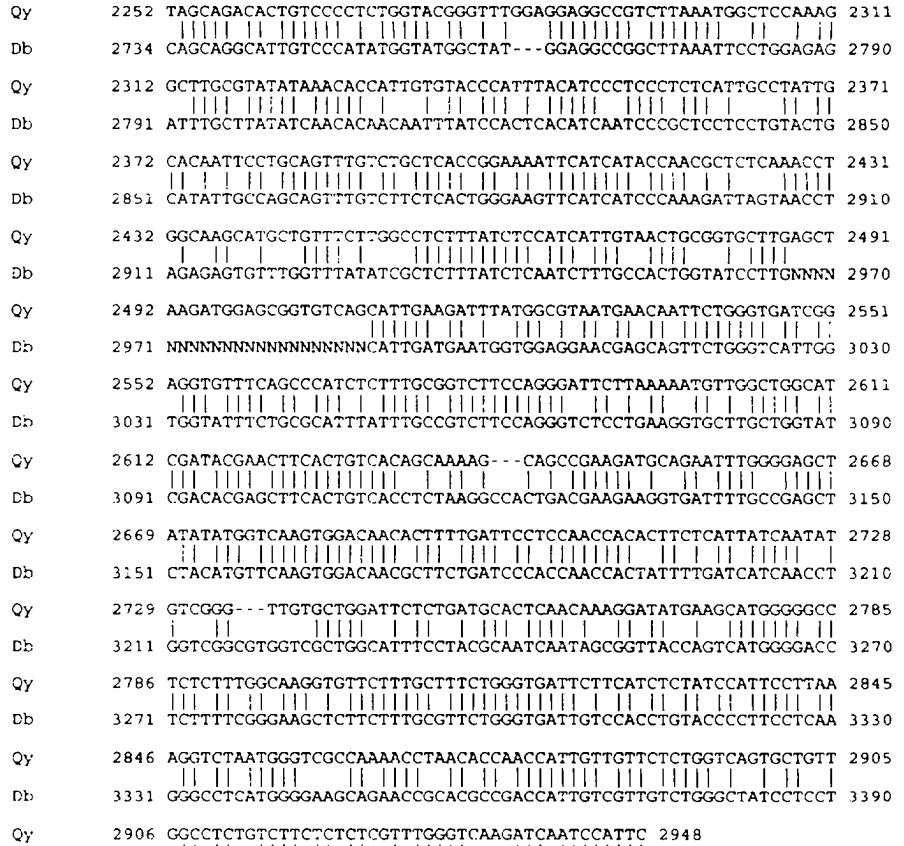
SULT 8
110079
CUS AY110079 3696 bp mRNA linear HTC 17-OCT 2002
FINITION Zea mays CL1164_1 mRNA sequence.
CESSION AY110079
RSION AY110079.1 GI:21214162
YWORDS HTC.
URCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
FERENCE 1 (bases 1 to 3696)
AUTHORS Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S., Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
TITLE Maize Mapping Project/DuPont Consensus Sequences for Design of Overgo Probes
JOURNAL Unpublished (2002)
FERENCE 2 (bases 1 to 3696)
AUTHORS Coe,E.H.

TITLE Direct Submissi
JOURNAL Submitted (25-A 2002) Maize Mapping Project, University of Missouri, Columbia, MO 65211, USA
COMMENT If you are interested in getting corresponding physical clones, these are publicly available from ZmDB and may be found by BLAST searching at MSU maizemap.org, ZmDB, www.zmdb.iastate.edu, TIGR, www.tigr.org, NCBI, www.ncbi.nlm.nih.gov. When the source of the maize cDNA sequences is either Virginia Walbot, Stanford or Pat Schnable, Iowa State, then clones may be requested from ZmDB: www.zmdb.iastate.edu.
FEATURES Location/Qualifiers
source 1..3696
/organism="Zea mays"
/mol_type="mRNA"
/db_xref="MaizeDB:430567"
/db_xref="taxon:4577"
/clone_lib="Maize Mapping Project/DuPont Consensus Library"
/note="This sequence is part of a project of BAT ... application of public ... this resource was ... by public support as part of a collaboration for the ... of BACs in conjunction with the Maize Mapping Project"
BASE COUNT 831 a 765 c 980 g 945 t 274 others
ORIGIN

Query Match 33.8%; Score 1065.8; DB 11; Length 3696;
Best Local Similarity 64.6%; Pred. No. 5.6e-138;
Matches 1579; Conservative 0; Mismatches 843; Indels 21; Gaps 8;







```
RESULT 9
BU895617
LOCUS       BU895617                571 bp    mRNA    linear   EST 17-OCT-2002
DEFINITION  X028D07 Populus wood cDNA library Populus tremula x Populus
            tremuloides cDNA 5 prime, mRNA sequence.
ACCESSION   BU895617
VERSION     BU895617.1  GI:24106824
KEYWORDS    EST.
SOURCE      Populus tremula x Populus tremuloides
  ORGANISM  Populus tremula x Populus tremuloides
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids I; Malpighiales; Salicaceae; Populus.
REFERENCE   1  (bases 1 to 571)
  AUTHORS   Unneberg,P., Bhalerao,R.R., Jansson,S. and Sterky,F.
  TITLE     The poplar tree transcriptome: Analysis of expressed sequence tags
            from multiple libraries
  JOURNAL   Unpublished
COMMENT     Contact: BHALERAO RUPALI R.
            Umea Plant Science Center
            Department of Plant Physiology
            University of Umea, 901 87 Umea, Sweden
            Tel: +46 90 786 5279
```

```
        Fax: +46 90 786 6676
        Email: rupali.bhalerao@plantphys.umu.se.
ATURES              Location/Qualifiers
   source           1. .571
                    /organism="Populus tremula x Populus tremuloides"
                    /mol_type="mRNA"
                    /db_xref="taxon:47664"
                    /tissue_type="wood"
                    /clone_lib="Populus wood cDNA library"
E COUNT      145 a    131 c    132 g    163 t
GIN

Query Match             17.2%;  Score 556.6;  DB 13;  Length 571;
Best Local Similarity   98.4%;  Pred. No. 2e-67;
Matches  562;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

      1827 GATCTCAATGCTGCCATATTTAATCTTACAGAGATTGATAATTATGACGAGCATGAAAGG 1886
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
         1 GATCTCAATGCTGCCATATTTAATCTTACAGAGATTGATAATTATGACGAGCATGAAAGG 60

      1887 TCAATGCTGATCTCCCAGTTGAGCTTTGAGAAAACTTTTGGCTTATCTTCTGTCTTCATT 1946
           |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
        61 TCAATGCTGATCTCCCAGTTGAGCTCTGAGAAAACTTTTGGCTTATCTTCTGTCTTCATT 120

      1947 GAGTCTACACTAATGGAGAATGGAGGAGTACCCGAGTCTGCCAACTCACCACCATTCATC 2006
           ||||||||||||||||||||||||||||||||||||||||||||||||||||| || |||||
       121 GAGTCTACACTAATGGAGAATGGAGGAGTACCCGAGTCTGCCAACTCACCAACACTCATC 180

      2007 AAGGAAGCGATTCAAGTCATCGGCTGTGGCTATGAAGAGAAGACTGAATGGGGAAAACAG 2066
           |||||||| ||||| ||||||||||||||||||||||||||||||||||||||||||| ||
       181 AAGGAAGCAATTCATGTCATCGGCTGTGGCTATGAAGAGAAGACTGAATGGGGAAAAGAG 240

      2067 ATTGGTTGGATATATGGGTCAGTCACTGAGGATATCTTAAGTGGCTTCAAGATGCACTGC 2126
           ||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
       241 ATTGGTTGGATCTATGGGTCAGTCACTGAGGATATCTTAAGTGGCTTCAAGATGCACTGC 300

      2127 CGAGGATGGAGATCAATTTACTGCATGCCCGTAAGGCCTGCATTCAAAGGATCTGCACCC 2186
           || ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
       301 CGCGGATGGAGATCAATTTACTGCATGCCCGTAAGGCCTGCGTTCAAAGGATCTGCACCC 360

      2187 ATCAACCTGTCTGATAGATTGCACCAGGTCCTCCGATGGGCTCTTGGTTCTGTGGAAATT 2246
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
       361 ATCAACCTGTCTGATAGATTGCACCAGGTCCTCCGATGGGCTCTTGGTTCTGTGGAAATT 420

      2247 TTCTTTAGCAGACACTGTCCCCTCTGGTACGGGTTTGGAGGAGGCCGTCTTAAATGGCTC 2306
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
       421 TTCTTTAGCAGACACTGTCCCCTCTGGTACGGGTTTGGAGGAGGCCGTCTTAAATGGCTC 480

      2307 CAAAGGCTTGCGTATATAAACACCATTGTGTACCCATTTACATCCCTCCCTCTCATTGCC 2366
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
       481 CAAAGGCTTGCGTATATAAACACCATTGTGTACCCATTTACATCCCTCCCTCTCATTGCC 540

      2367 TATTGCACAATTCCTGCAGTTTGTCTGCTCA 2397
           |||||||||||||||||||||||||||||||
       541 TATTGCACAATTCCTGCAGTTTGTCTGCTCA 571


SULT 10
867361
CUS       BQ867361                731 bp    mRNA    linear   EST 14-AUG-2002
FINITION  QGD10J04.yg.ab1 QG_ABCDI lettuce salinas Lactuca sativa cDNA clone
          QGD10J04, mRNA sequence.
CESSION   BQ867361
RSION     BQ867361.1  GI:22253118
YWORDS    EST.
URCE      Lactuca sativa
ORGANISM  Lactuca sativa
          Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
          Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
          asterids; campanulids; Asterales; Asteraceae; Cichorioideae;
          Cichorieae; Lactuca.
FERENCE   1  (bases 1 to 731)
AUTHORS   Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,
          Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegle,J., Ellison
          ,P., Kolkman,J., Slabaugh,M.S., Livingston,K., Zhou,Y., Lai,Z.,
          Church,S., Jacks,T.L. and Bradford,K.
  TITLE     Lettuce and Sunflower ESTs from the Compositae Genome Project
            http://compgenomics.ucdavis.edu/
  JOURNAL   Unpublished
COMMENT     Contact: Alexander Kozik [R.W.Michelmore]
            Department of Vegetable Crops, R.W.Michelmore Lab
            University of California at Davis (UCD)
            Asmundson Hall, 1 Shields Ave, Davis, CA 95616, USA
            Tel: 1-(530)-752-1742
            Fax: 1-(530)-752-4659
            Email: akozik@ucdavis.edu [rwmichelmore@ucdavis.edu]
            belongs to contig CA_Contig5167, see http://cgpdb.ucdavis.edu/
            for details.
            Plate: QGD10  row: J  column: 04.
FEATURES              Location/Qualifiers
     source           1. .731
                      /organism="Lactuca sativa"
                      /mol_type="mRNA"
                      /cultivar="Salinas"
                      /db_xref="taxon:4236"
                      /sex="hermaphrodite"
                      /tissue_type="flowers pre-fertilized"
                      /note="Vector: pBR-cDNAstAB; The library was constructed
                      from 10 different sources of RNA from a single genotype.
                      Separate cDNAs were generated using primers that
                      incorporated unique 5' and 3' tags to distinguish each
                      source of RNA. cDNAs were then pooled, size-fractionated,
                      directionally cloned into a custom medium copy vector and
                      transformations made with four size classes to minimize
                      size bias. Details of each source of RNA and library
                      construction can be obtained at http://cgpdb.ucdavis.edu/
                      TAG_LIB QG ABCDI lettuce salinas
                      TAG_TISSUE=flowers pre-fertilized
                      TAG_SEQ GTTTGAGGGG"
BASE COUNT      209 a    142 c    160 g    219 t
ORIGIN

  Query Match             14.8%;  Score 481.2;  DB 13;  Length 731;
  Best Local Similarity   78.4%;  Pred. No. 4.1e-57;
  Matches  573;  Conservative    0;  Mismatches  158;  Indels    3;  Gaps    3;

Qy        536 GGAGGCGGCTTTCAATTTTTTATGCAATTGCACGCAACAAGCTCACACCATACAGAGCAGT 595
              | | || ..||| |  || .|||'' | | | ;|||||| .| ||'|  ||
Db          1 GCAACCACTTTCATAAATAGTTTCAATTGCAAAGAGCAAATCACACCTTATAGAGTTGT 60

Qy        596 GATCATTATGGCGACTGGTCATTCTTGGGCCTCTTCTTCGCACTTCAGAATAACAAATCCTGT 655
              || ||  |||| .  || ||| |  | |||| | || |||||||||||| || ||
Db         61 AATTATAGTGCGATTGATCATTTTGGGACTTTTCTTTCATTACAGAATAACAAACCCAGT 120

Qy        656 CGATAGTGCCTTTGAGCTGTGGCTTACTTCTGTCATATGTGAGATCTGGTTTGCATTTTC 715
              || ||| | | ||| | |: || .|||| . || ||||||||||||||| ||||
Db        121 TGAAAGTTCTTATGTTTATGGCTAACTTCTGTATTTGTGAGATCTGGTTTGCTATTTC 180

Qy        716 TTGGGTCTTGGATCAGTTCCCCAAGTGGAATCTTGTCAATAGAGAAACGTATATCGAAAG 775
              |||||| ||||,|| .|||.| || ||  |||||||  || | | |||
Db        181 TTGGGTCTTGGATTGATTCCCCAAATGGTGTCTCATCAATAGAATTACTTTCACCGATGA 240

Qy        776 GCTGTCGGCAAGGTATGAAAGAGAGGGTGAGCTTCTCAGCTTGCTGGTGTGGATTTTTT 835
              || || ||||||| ||||||| |||| ||||| || |||| | ||| | | ||| |
Db        241 ACTATCCGCAAGGTATGAAAGAGAAGGTGAGCTTCACAACTTGCGGCTGTTGACTTCTT 300

Qy        836 CGTGAGTACTGTTGATCGCTGAAGGAACCGCCATTGATCACTGCCAATACAGTCCTTTC 895
              .|||| .| || | :. ||| || || .. | .|||| ||||||| || | .||
Db        301 TGTGAGCACAGTTGATCCTTTGAAAGAGCCTCCTCTAATCACAGCCAATACTGTACTTTC 360

Qy        896 CATCCTTGCTGTGGACTATCCCGTCGATAAAGTCTCCTGCTACGTGTCTGATGATGGTGC 955
              || || |||||  | || ||||| ||:|||||| || || || || | |||.|||||
Db        361 AATCCTAGCTGTTGATTACCCCGTTGATAAAGTCTCATGTTATGTCTCGGATGATGGTGC 420

Qy        956 AGCTATGCTTTCATTTGAATCTCTTGTAGAAACAGCTGAGTTTGCAAGGAAGTGGGTTCC 1015
```

```
Db        421 TGCAATGCTTTCGTTTGAATCGCTTGTTGAAACCGCGGAGTTTGCAAGGAAGTGGGTCGC 480

Qy       1016 GTTCTGCAAAAAATTCTCAATTGAACGAAGAGCACCGGAGTTTACTTCTCACAGAAAAT 1075

Db        481 GTTTTGTAAACGGTTTTCAATTGAGCCACGTGCACCTGAATTTACTTCTCCCAAAAGAT 540

Qy       1076 TGATTACTTGAAAGACAAGGTTCAACTTCTTTCGTGAAAAACGTAGAGCAATGAAAAG 1135

Db        541 CGATTACTTGAAAGATAAAGTCTAACTTCATTTGTGAAAAACGTAGAGCAATGAAAACG 600

Qy       1136 GGATTATGAAGAGTACAAAGTCCGAGTTAATGCCCTGGTATAAAGGCTCAGAAAACACC 1195

Db        601 GGATTATGAGGAGTATAAAGTGAGGTTAATGTTTAGTCTTAAAGCTCAGAAAACCCC 660

Qy       1196 TGAAGAAGGATGGACTATGCAAGATGGAACACCTTGGCCTCGGAATAACACACGTGATCA 1255

Db        661 GGAAGAAGGATGGACCATGCAAGATGGAACGCGTGGCCGCGGAATAACACACGTGACCA 720

Qy       1256 CCCTGG 1261

Db        721 CCCGGG 726


RESULT 11
BU894292
LOCUS       BU894292                 648 bp    mRNA    linear   EST 17-OCT-2002
DEFINITION  X005H01 Populus wood cDNA library Populus tremula x Populus
            tremuloides cDNA 5 prime, mRNA sequence.
ACCESSION   BU894292
VERSION     BU894292.1  GI:24105957
KEYWORDS    EST
SOURCE      Populus tremula x Populus tremuloides
  ORGANISM  Populus tremula x Populus tremuloides
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids I; Malpighiales; Salicaceae; Populus.
REFERENCE   1  (bases 1 to 648)
  AUTHORS   Unneberg,P., Bhalerao,R.P., Jansson,S. and Sterky,F.
  TITLE     The poplar tree transcriptome: Analysis of expressed sequence tags
            from multiple libraries
  JOURNAL   Unpublished
COMMENT     Contact: BHALERAO RUPALI P
            Umea Plant Science Center
            Department of Plant Physiology
            University of Umea, 901 87 Umea  Sweden
            Tel: +46 90 786 5279
            Fax: +46 90 786 6676
            Email: rupali.bhalerao@upsc.[illegible]
FEATURES             Location/Qualifiers
     source          1..648
                     /organism="Populus tremula x Populus tremuloides"
                     /mol_type="mRNA"
                     /db_xref="taxon:47664"
                     /tissue_type="wood"
                     /clone_lib="Populus wood cDNA library"
BASE COUNT      173 a    121 c    141 g    213 t
ORIGIN

  Query Match           14.7%;  Score 475.2;  DB 14;  Length 648;
  Best Local Similarity   89.1%;  Pred. No. 5.1e-56;
  Matches  542;  Conservative    0;  Mismatches   48;  Indels   18;  Gaps    2;

Qy       2627 TGTCACAGCAAAAGCAGCCGAAGATGCAAGAATTTGGGGAGATATATATGGTCAAGTGGAC 2686

Db          2 TGTCACAGCAAAAGCAGCCGAAGATGCAAGAATTTGGGGAGATATATATGGTCAAGTGGAC 61

Qy       2687 AACACTTTTGATTCCTCCAACAATACTCTCTATATATCAATCTCTCTTGTTGCTGG 2743

Db         62 AACACTTTTGATTCCTCCAACAATACTCTCTATATATCAATCTCTCTTGTTGCTGG 121

Qy       2744 ATTCTCTGATGCACTCAACAAAGGATATGAAGCATGGGGGCCTCTCTTTGGCAAGGTGTT 2803

Db        122 ATTCTCTGATGCACTCAACAAAGGATATGAAGCATGGGGGCCTCTCTTTGGCAAGGTGTT 181

Qy       2804 CTTTGCTTTCTGGGTGATTCTTCATCTCTATCCATTCCTTAAAGGTCTAATGGGTCGCCA 2863

Db        182 CTTTGCTTTCTGGGTGATTCTTCATCTCTATCCATTCCTTAAAGGTCTAATGGGTCGCCA 241

Qy       2864 AAACCTAACACCAACCATTGTTGTTCTCTGGTCAGTGCTGTTGGCCTCTGTCTTCTCTCT 2923

Db        242 AAACCGAACACCAACCATTGTTGTTCTCTGGTCAGTGCTGTTGGCCTCTGTCTTCTCTCT 301

Qy       2924 CGTTTGGGTCAAGATCAATCCATTCGTTAACAAAGTTGATAACACCTTGGTTGCGGAGAC 2983

Db        302 TGTTTGGGTCAAGATTAATCCATTCGTTAACAAAGTTGATAACACCTTGGTTGCGGAGAC 361

Qy       2984 CTGCATTTCCATTGATTGCTGAGCTACCTCCAATAAGTCTCTCCCAGTATTTTGGGGTTA 3043

Db        362 CTGCATTTCCATTGATTGCTGAGCTACCTCCAATAAG----------TGGGGTTA 406

Qy       3044 CAAAACCTTTGGGAATTGGAATATGATCCTCGTTGTAGTTTCCCTCAAGAAAGCACATAT 3103

Db        407 CAAAACCTTTGGGAATTGGAATATGATCCTCGTTGTAGTTTCCCTCAAGAAAGCACATAT 466

Qy       3104 CGCTGTCAGTATTTAAATGAACTGCAAGATGATTGTTCTCTATGAAGTTTTGAACAGTTT 3163

Db        467 CGCTGTCAGTATTTAAATGGACTGCAAGATGATTGTTCTCTATGAAGTTTTGAACAGTTT 526

Qy       3164 GAAATGATATTATGTTAAAATACAGGTTTTGATTGTGTTGAAAAAAAAAAAGAAAAAAAA 3223

Db        527 GAAATGCTCTTTTTGCTGAGCGAAAGGTCTAAGAATTGTTGTGACGAGATATTATGTAAA 586

Qy       3224 AAAAAAAA 3231

Db        587 AATATAGA 594


RESULT 12
CB648832
LOCUS       CB648832                 800 bp    mRNA    linear   EST 08-APR-2003
DEFINITION  OSJNEb12G21.f OSJNEb Oryza sativa (japonica cultivar-group) cDNA
            clone OSJNEb12G21 5', mRNA sequence.
ACCESSION   CB648832
VERSION     CB648832.1  GI:29643825
KEYWORDS    EST.
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 800)
  AUTHORS   Jantasuriyarat,C., Li,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
            Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.
  TITLE     Large-scale identification of ESTs involved in the interaction
            between rice and Magnaporthe grisea
  JOURNAL   Unpublished
COMMENT     Contact: Rod Wing
            Arizona Genomics Institute
            University of Arizona
            Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
            85721-0088, USA
            Tel: 520 626 3967
            Fax: 520 621 9288
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: gta aaa cga cgg cca gtg
            BACKWARD: gga aac agc tat gac cat g
            Plate: 12  row: G  column: 21
            Seq primer: gta aaa cga cgg cca gtg.
FEATURES             Location/Qualifiers
     source          1..800
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
```

/clone="OSJNEb12G21"
/tissue_type="Leaf"
/dev_stage="3 week"
/lab_host="DH10B"
/clone_lib="OSJNEb"
/note="Vector: pBluescript II KS +; Site 1: EcoRI; Site 2:
XhoI; 24 hrs after innoculation with Rice Blast (Che 86061
)"

SE COUNT 211 a 176 c 194 g 219 t
GIN

Query Match 14.6%; Score 470.6; DB 14; Length 800;
Best Local Similarity 75.7%; Pred. No. 1.3e-55;
Matches 597; Conservative 0; Mismatches 189; Indels 5; Gaps 1;

```
473 AACTGAACCAGCTCAAGTTCCTACAGAACAGCAGATGGAAGAGAAACCGTCTGCAGAGGC 532
    ||  |  | |||||  |  |  |  |||||||   |||||  || |||   |   ||||  |
  2 AAAAGCTCAAGCTCCACCTGTTGAAGAACAGATTATGGATGAAAAAGACTTGACAGATGC 61

533 TTCGGAGCCGCTTTCAATTGTTTATCCAATTCCACGCAACAAGCTCACACCATACAGAGC 592
    |  ||||| |||||      |  |||  |   |  |||||||||||  | | ||
 62 ATATGAGCCACTTTCTCGGATCATTCCTATATCGAAGAACAAGCTCACACTTACAGGAC 121

593 AGTGATCATTATGCGACTGGTCATTCTGGGCCTCTTCTTCCACTTCAGAATAACAAATCC 652
    ||| ||||||||||| |||||  |||||| |||||||| |||| | | || || ||| |
122 AGTAATCATTATGCGTCTGGTAGTTCTGGGGCTCTTCTTTCACTACCGTATCACCAACCC 181

653 TGTCGATAGTGCCTTTGGCCTGTGGCTTACTTCTGTCATATGTGAGATCTGGTTTGCATT 712
    |||  | || |||||||| || ||| | || || |||||||||||||||||| ||| |||
182 TGTTTACAGCGCCTTTGGTCTCTGGATGACATCAGTCATATGTGAGATCTGGTTTGGATT 241

713 TTCTTGGGTGTTGGATCAGTTCCCCAAGTGGAATCCTGTCAATAGAGAAACGTATATCGA 772
    || ||| |  |||||||||||||||||||  |||| |||||  | |||||  ||| | ||
242 CTCCTGGATCCTGGATCAGTTCCCCAAGTGGTGTCCTATCAATCGCGAAACTTATGTTGA 301

773 AAGGCTGTCGGCAAGGTATGAAAGAGAGGGTGAGCCTTCTCAGCTTGCTGGTGTGGATTT 832
    ||||||   ||| | ||||   |||| || ||   ||||  | | |  ||| |||||
302 TAGGCTGATTGCACGATATG---GAGATGGAGAAGATTCTGGGTTGGCACCTGTAGATTT 358

833 TTTCGTGAGTACTGTTGATCCGCTGAAGGAACCGCCATTGATCACTGCCAATACAGTCCT 892
    || || ||||| || ||||  |||| || || ||  | |||||||| |||| || ||
359 CTTTGTCAGTACAGTGGATCCATTGAAAGAGCCTCCTCTAATCACTGCGAATACTGTTCT 418

893 TTCCATCCTTGCTGTGGACTATCCCGTCGATAAAGTCTCCTGCTACGTGTCTGATGATGG 952
    ||||| ||||| |||||||||||  || || ||  | || || ||||||| || ||||
419 GTCCATTCTTGCCGTGGACTATCCAGTTGAGAAGATTTCGTGTTACGTGTCAGACGATGG 478

953 TGCAGCTATGCTTTCATTTGAATCTCTTGTAGAAACAGCTGAGTTTGCAAGGAAGTGGGT 1012
    | | ||||||||  |||||||||| |||| |||||||| || ||||| ||.|  ||||
479 TTCGGCTATGCTCACATTTGAATCACTTGCGGAAACAGCAGAATTTGCTAGGAGATGGGT 538

1013 TCCGTTCTGCAAAAAATTCTCAATTGAACCAAGAGCACCGGAGTTTTACTTCTCACAGAA 1072
     |:| || ||||| || | ||| ||||| ||| | || || |||||  |||||||||||.:|
539  TCCATTTTGCAAGAAGTACTCCATTGAGCCACGTGCCCCTGAGTTCTACTTCTCACAGAA 598

1073 AATTGATTACTTGAAAGACAAGGTTCAACCTTCTTTCGTGAAAGAACGTAGAGCAATGAA 1132
     |||||  |||||||| |||||| | || || |||||||  || || ||  | |:  ||||
599  GATTGACTACTTGAAGGACAAGATACACCCATCTTTCGTCAAGGAGCGTCGGGCTATGAA 658

1133 AAGGGATTATGAAGAGTACAAAGTCCGAGTTAATGCCCTGGTAGCAAAGGCTCAGAAAAC 1192
     |||||||| ||||| ||||| ||  |  | |||||  |||| || ||||||||||| ||
659  GAGGGATTACGAAGAATACAAGGTGAGGATAAATGCTTTGGTTGCTAAGGCTCAGAAGAC 718

1193 ACCTGAAGAAGGATGGACTATGCAAGATGGAACACCTTGGCCTGGGAATAACACACGTGA 1252
     |||||| ||||| |||| |||||||||||| ||||| |||||||||| || | |||||
719  ACCTGAGGAAGGCTGGATTATGCAAGATGGGACACCATGGCCTGGGAACAACCCTCGTGA 778

1253 TCACCCTGG 1261
     |||||||||
779  CCACCCTGG 787
```

SULT 13

BI128887
LOCUS BI128887 484 bp mRNA linear EST 31-DEC-2001
DEFINITION G083P04Y Populus tremula x Populus tremuloides cDNA library Populus tremula x Populus tremuloides cDNA mRNA sequence.
ACCESSION BI128887
VERSION BI128887.1 GI:17867
KEYWORDS EST.
SOURCE Populus tremula x Populus tremuloides
ORGANISM Populus tremula x Populus tremuloides
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids I; Malpighiales; Salicaceae; Populus.
REFERENCE 1 (bases 1 to 484)
AUTHORS Hertzberg,M., Aspeborg,H., Sundberg,B., Bjorkbacka,H., Biltonen,T., Karlsson,J., Moritz,T., Rohrafo,B., Bahlerao,R., Jansson,S., Nilsson,O., Sitbon,F., Welander,M., Olsen,J., Ohlen,M., Sandberg,G. and Lundeberg,J.
TITLE Gene expression in poplar
JOURNAL Unpublished
COMMENT Contact: Bo Oxelman
Department of Systematic Botany
Royal Institute of Technology
Tekniksvagen 14, 10044 Stockholm
Tel: 46 8 7216562
Fax: 46 8 21545
Email: mikerabbit@hem.kth.se
FEATURES Location/Qualifiers
source 1..484
/organism="Populus tremula x Populus tremuloides"
/mol_type="mRNA"
/db_xref="taxon:47664"
/clone_lib="Populus tremula cDNA library"
/note="vector: pCMVSPORT6"
BASE COUNT 122 a 115 c 101 g 146 t
ORIGIN

Query Match 14.1%; Score 455.8; DB 12; Length 484;
Best Local Similarity 92.4%; Pred. No. 1.2e-53;
Matches 474; Conservative 0; Mismatches 7; Indels 5; Gaps 1;

```
Qy   2290 GCCGTCTTAAATGG-TCCAAAGGCTTC-GTATATAAACACCATTGTGTACCCATTTACAT 2348
          |:|||:|||.:|||  ||||.|||.:|||  |||||'|||||.'||||:'||||:|
Db      1 GCCGTCTTAAATGGGTCCAAAGGCTTCGGTATATAAACACCATTGTGTACCCATTTACAT 60

Qy   2350 CCCTCCCTCCTCATT-TTATTGCAAAATTCCTGCAGTTTGTCTGCTCACCGGAAAATTCA 2408
          ||||||.|||.|:|.|||.|||  |||.'|||||||.|||  |||||||||||||||||
Db      61 CCCTCCCTCCTCATTGTATTGCAAAATTCCTGCAGTTTGTCTGCTCACCGGAAAATTCA 120

Qy   2410 TCATACCAA-GCTC-AAAGCTTGGCAAGCATGGTGTTTCTTGGCCTCTTTATCTCCATCA 2468
          ||||||||| |||. ||| |||  ||| |||:|||||.|||||||||||||||||||||
Db     121 TCATACCAAAGCTCTAAAGCTTGGCAAGCATGGTGTTTCTTGGCCTCTTTATCTCCATCA 180

Qy   2470 TTGTAACTG-GGT-TGAGCTAACATGAGCGGTGTCAGCATTGAAGATTTATGGCGTA 2528
          |||:|||  ||| .||. |||| |||'|||||'|||||'|||:'|||||||:|||||
Db     181 TTGTAACTGAGGTATGAGCTTAACATGAGCGGTGTCAGCATTGAAGATTTATGGCGTA 240

Qy   2530 ATGAACAATTCTGG-GATGGAGGTTCCATCCCATCTCTTTGCGGTCTTCCAGGGAT 2588
          ||||.:'||.:||  ||| ||| .'||  '||||:'||||.'|||'|||||  .'||
Db     241 ATGAACAATTCTGGTGATGGAGGTTCCATCCCATCTCTTTGCGGTCTTCCAGGGAT 300

Qy   2589 TCTTAAAAAATGTTG-TGGCATGGATAAGAACTTCACTGTCACAGCAAAAGCAGCCGAAG 2647
          ||||: |||  |||  |||||||||||.|.|| ||||||||||||  ||||||.|||
Db     301 TCTTAAAAAAGGTTGTGGCATGGATAAGAACTTCACTGTCACAGCAAAAGCAGCCGAAG 360

Qy   2650 ATGCAGAATTTGGG-AGCTATATATAGTCAAGTGGACAACACTTTTGATTCCTCCAACCA 2708
          |||||  |||  |||  |||' |||  |||| |||.:|||||||||:|||||:|| |
Db     361 ATGCAGAATTTGGGTAGCTATATATAGTCAAGTGGACAACACTTTTGATTCCTCCAACAA 420

Qy   2710 CACTTCTCATTATCAATATGTCGGG-TTGTGCTGGATTCTCTGATGCACTCAACAAAG 2766
          ||||||.|||.|:|.|  '|    |.' ||||' ||||' ||||'  ||||.:|
Db     421 CACTTCTCATTATCAATATTGTCGGTGTTGTTGCTGGATTCTCTGATGCACTCAACAAAG 480

Qy   2767 GATA 2770
```

```
                ||.|
Db        481 GATA 484


RESULT 14
BQ969585
LOCUS       BQ969585                735 bp    mRNA    linear   EST 21-AUG-2002
DEFINITION  QHB38P07.yg.ab1 QH_ABCDI sunflower RHA801 Helianthus annuus cDNA
            clone QHB38P07, mRNA sequence.
ACCESSION   BQ969585
VERSION     BQ969585.1  GI:22387106
KEYWORDS    EST.
SOURCE      Helianthus annuus (common sunflower)
  ORGANISM  Helianthus annuus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            asterids; campanulids; Asterales; Asteraceae; Asteroideae;
            Heliantheae; Helianthus.
REFERENCE   1  (bases 1 to 735)
  AUTHORS   Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,
            Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegle,J., Ellison
            ,P., Kolkman,J., Slabaugh,M.S., Livingston,K., Zhou,Y., Lai,Z.,
            Church,S., Jackson,L. and Bradford,K.
  TITLE     Lettuce and Sunflower ESTs from the Compositae Genome Project
            http://compgenomics.ucdavis.edu/
  JOURNAL   Unpublished
COMMENT     Contact: Alexander Kozik [R.W.Michelmore]
            Department of Vegetable Crops, R.W.Michelmore Lab
            University of California at Davis (UCD)
            Asmundson Hall, UCD, Davis, CA 95616, USA
            Tel: 1 (530)-742-1742
            Fax: 1 (530)-752-9659
            Email: akozik@atgc.org (m.chelmorevegmail.ucdavis.edu)
            singleton, see http://cgpdb.ucdavis.edu/ for details.
            Plate: QHB38  row: P  column: 07
FEATURES             Location/Qualifiers
     source          1. .735
                     /organism="Helianthus annuus"
                     /mol_type="mRNA"
                     /cultivar="RHA801"
                     /db_xref="taxon:4232"
                     /clone="QHB38P07"
                     /lab_host="E.coli"
                     /clone_lib="QH ABCDI sunflower RHA801"
                     /note="Vector: pBR(CONAS)1AB; The library was constructed
                     from 11 different stages of RNA from a single genotype.
                     Separate cDNAs were primed from [illegible] primers that
                     incorporated [illegible] to distinguish each
                     source of RNA. cDNA was [illegible], size fractionated,
                     directionally [illegible] into [illegible] copy vector and
                     transformations made with four [illegible] classes to minimize
                     size bias. Details of each source of RNA and library
                     construction can be obtained at http://cgpdb.ucdavis.edu/
                     TAG_LIB.QH_ABCDI sunflower RHA801
                     TAG_TISSUE=chemical induction
                     TAG_SEQ=TGTAGCCGGG"
BASE COUNT      170 a    159 c    188 g    218 t
ORIGIN

  Query Match             14.1%;  Score 455.6;  DB 1;  Length 735;
  Best Local Similarity   76.3%;  Pred. No. 1.5e-53;
  Matches  560;  Conservative    0;  Mismatches  174;  Indels    0;  Gaps    0;

Qy       1983 TCTGCCAACTCACCACCATTCATCAAGGAAGCGATTCAAGTTATCGGCTGTGGCTATGAA 2042
             |:|||:||| |  |  |  | || |  |||   ||.|: |:  || | |||| ||||||
Db          1 TCTGCAAACCCTGCAACTATGATCAATGAAGCAATTCATTTATCAGTTGTGGGTATGAA 60

Qy       2043 GAGAAGACTGAATGGGGAAAAACAGACTGGTTGGATATATGTTCAGTCACTGAGGATATC 2102
             ||||| || | |||||| ||| ||:| ||  ||| ||||  | |  || |:|||||||
Db          61 GAGAAAACCGCATGGGGTAAAGAGATTGGCTGGATTTATTTATCGGTTACAGAGGATATC 120

Qy       2103 TTAAGTGGCTTCAAGATGCACTGTCCAAGAAGAGATCAATTTACTGCATGCCCGTAAGG 2162
             ||||   | |||| ||||  ||    ||  |||'  :  ||| |||   ||| |

Db        121 TTAACGGGATTCAAGATGCATTGTAGGGGATGGAGATCGATATACTGCATGCCAGTAAGA 180

Qy       2163 CCTGCATTCAAAGGATCTGCACCCATCAACCTGTCTGATAGATTGCACCAGGTCCTCCGA 2222
             || ||||| ||||| || || || |||||| | || ||| | ||||||||||  |  |
Db        181 CCGGCATTTAAAGGTTCGGCTCCTATCAACTTATCAGATCGGTTGCACCAGGTTTTGAGG 240

Qy       2223 TGGGCTCTTGGTTCTGTGGAAATTTTCTTTAGCAGACACTGTCCCCTCTGGTACGGGTTT 2282
             ||||| ||||| || || || || || || || || || || ||  | ||||| || |
Db        241 TGGGCCCTTGGATCGGTTGAGATCTTTTTGAGTAGGCATTGCCCGTTGTGGTATGGTTGG 300

Qy       2283 GGAGGAGGCCGTCTTAAATGGCTCCAAAGGCTTGCGTATATAAACACCATTGTGTACCCA 2342
             || || ||||| ||||| | ||| ||| | |||||||||||| |||||||| || || ||
Db        301 GGTGGCGGCCGCCTTAAGTTGCTTCAACGACTTGCGTATATTAACACCATCGTTTATCCG 360

Qy       2343 TTTACATCCCTCCCTCTCATTGCCTATTGCACAATTCCTGCAGTTTGTCTGCTCACCGGA 2402
             || ||||| |||||||| |||| || |||||| | || ||  | || || || ||||||
Db        361 TTCACATCTCTCCCTCTTGTTGCATACTGCACATTGCCCGCCATCTGCCTTCTGACCGGA 420

Qy       2403 AAATTCATCATACCAACGCTCTCAAACCTGGCAAGCATGCTGTTTCTTGGCCTCTTTATC 2462
             |||||||||||| |||||||| || || || |||     |   ||||||| ||||||||
Db        421 AAATTCATCATCCCAACGCTTTCCAATCTAGCAGCGGTTTGGTTTCTTTGCCTCTTTTTG 480

Qy       2463 TCCATCATTGTAACTGCGGTGCTTGAGCTAAGATGGAGCGGTGTCAGCATTGAAGATTTA 2522
             || || |||  |||    || |||||  |  |||||||||||||||||| |||||  |
Db        481 TCGATTATTACAACCAGTGTTCTTGAAATTCGATGGAGCGGTGTCAGCATCGAAGAGCTT 540

Qy       2523 TGGCGTAATGAACAATTCTGGGTGATCGGAGGTGTTTCAGCCCATCTCTTTGCGGTCTTC 2582
             ||| | || || ||||| ||||| ||||| || || ||||| ||||| || ||
Db        541 TGGAGAAACGAGCAATTTTGGGTCATCGGTGGTGTCTCCGCTCATCTGTTTGCCGTGTTT 600

Qy       2583 CAGGGATTCTTAAAAATGTTGGCTGGCATCGATACGAACTTCACTGTCACAGCAAAAGCA 2642
             || |||||| | |||||||||||| ||  | ||||| |||||||| || || ||||||||
Db        601 CAAGGATTCCTGAAAATGTTGGCGGGTGTGGATACCAACTTCACAGTTACCGCAAAAGCA 660

Qy       2643 GCCGAAGATGCAGAATTTGGGGAGCTATATATGGTCAAGTGGACAACACTTTTGATTCCT 2702
             || || |||   || || |||||||| |||||| |||| ||||||||  |||||||  |||
Db        661 GCTGATGATCAGGAGTTCGGGGAGCTTTATATGATCAAATGGACAACGGTTTTGATCCCT 720

Qy       2703 CCAACCACACTTCT 2716
             || || || || ||
Db        721 CCGACGACCCTGCT 734


RESULT 15
BF635224
LOCUS       BF635224                668 bp    mRNA    linear   EST 19-DEC-2000
DEFINITION  NF061H04DT1F1043 Drought Medicago truncatula cDNA clone NF061H04DT
            5', mRNA sequence.
ACCESSION   BF635224
VERSION     BF635224.1  GI:11899382
KEYWORDS    EST.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 668)
  AUTHORS   Torrez-Jerez,I., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell,C.J.,
            Flores,H.R., Inman,J.T., Weller,J.W. and May,G.D.
  TITLE     Expressed Sequence Tags from the Samuel Roberts Noble Foundation
            Medicago truncatula drought library
  JOURNAL   Unpublished
COMMENT     Contact: May GD
            Plant Biology Division
            The Samuel Roberts Noble Foundation
            2510 Sam Noble Parkway, Ardmore, OK 73402, USA
            Tel: 580 221 7391
            Fax: 580 221 7380
            Email: gdmay@noble.org
            Insert Length: 668   Std Error: 0.00
            Plate: 061  row: H  column: 04
            Seq primer: TCACACAGGAAACAGCTATGAC.
```

```
ATURES              Location/Qualifiers
   source           1. .668
                    /organism="Medicago truncatula"
                    /mol_type="mRNA"
                    /db_xref="taxon:3880"
                    /clone="NF061H04DT"
                    /tissue_type="Plantlets"
                    /dev_stage="Pooled timepoints"
                    /clone_lib="Drought"
                    /note="Vector: Lambda Zap; Contains a mixture of entire
                    plantlets harvested in a series of days-post-watering
                    timepoints. "
SE COUNT      157 a    140 c    136 g    235 t
IGIN

Query Match             14.0%;  Score 453.6;  DB 10;  Length 668;
Best Local Similarity   81.0%;  Pred. No. 3e-53;
Matches  541;  Conservative    0;  Mismatches  124;  Indels    3;  Gaps    1;

      2274 TACGGGTTTGGAGGAGGCCGTCTTAAATGGCTCCAAAGGCTTGCGTATATAAACACCATT 2333
           ||||| |||| || || ||||| ||| ||| || || || || |||||||| |||
         1 TACGGTTTTGCTGGCGGTCGTCTGAAACTGCTGCAGAGACTGGCTTATATAAACACCATC 60

      2334 GTGTACCCATTTACATCCCTCCCTCTCATTGCCTATTGCACAATTCCTGCAGTTTGTCTG 2393
           || ||||| || |||||||| |||||| |||| ||||| ||  | |||||| |||| ||
           61 GTTTACCCTTTCACATCCCTTCCTCTCGTTGCTTATTGTACTTTGCCTGCAATTTGCCTA 120

      2394 CTCACCGGAAAATTCATCATACCAACGCTCTCAAACCTGGCAAGCATGCTGTTTCTTGGC 2453
           ||||| |||||||||||||||||||||| |||||| | |||||    || ||||||||
           121 CTCACAGGAAAATTCATCATACCAACGCTATCAAACATTGCAAGTGCACTCTTTCTTGGA 180

      2454 CTCTTTATCTCCATCATTGTAACTGCGGTGCTTGAGCTAAGATGGAGCGGTGTCAGCATT 2513
           || || |||||||| ||  | ||    || |||||||| || ||||||||||| | |||
           181 CTTTTCATCTCCATTATATTGACCAGTGTACTTGAGCTGAGGTGGAGCGGTGTTACCATA 240

      2514 GAAGATTTATGGCGTAATGAACAATTCTGGGTGATCGGAGGTGTTTCAGCCCATCTCTTT 2573
           |||||||||||||||| |||||||||||||||| |||||||||| || |||| |||
           241 GAAGATTTATGGCGTAACGAACAATTCTGGGTGATTGGAGGTGTTTCGGCTCATCTTTTT 300

      2574 GCGGTCTTCCAGGGATTCTTAAAAATGTTGGCTGGCATCGATACGAACTTCACTGTCACA 2633
           || || ||||| || ||| |||| ||||| || ||  | ||||| ||||||||| ||||
           301 GCAGTGTTCCAAGGTTTCCTAAAGATGTTAGCAGGTGTTGATACAAACTTCACTGTCACA 360

      2634 GCAAAAGCAGCCGAAGATGCAGAATTTGGGGAGCTATATATGGTCAAGTGGACAACACTT 2693
           ||| |||| || || |||||||| ||||| || |||||||| |||| ||||||||||
           361 GCAAAAGCTGCTGATGATGCAGAGTTTGGTGAACTATATATGATCAAATGGACAACACTC 420

      2694 TTGATTCCTCCAACCACACTTCTCATTATCAATATGTCGGGTTGT---GCTGGATTCTCT 2750
           || |||||||||| |  ||| ||||||| || |  ||| |  ||||||| |||
           421 TTAATTCCTCCAACAAGTCTTATCATTATTAACCTTGTTGGTGTTGTAGCTGGATTTTCT 480

      2751 GATGCACTCAACAAAGGATATGAAGCATGGGGGCCTCTCTTTGGCAAGGTGTTCTTTGCT 2810
           |||||||| ||    |||||||| | ||||| || ||| |||| || || ||||| ||
           481 GATGCACTTAATGGTGGATATGAATCTTGGGGACCACTCATTGGAAAAGTTTTCTTCGCA 540

      2811 TTCTGGGTGATTCTTCATCTCTATCCATTCCTTAAAGGTCTAATGGGTCGCCAAAACCTA 2870
           || |||||||| ||||||| |||||||||||||| ||||| ||||||||||||||||
           541 TTTTGGGTGATTTTTCATCTTTATCCATTCCTTAAGGGTCTCATGGGTCGCCAAAACCGT 600

      2871 ACACCAACCATTGTTGTTCTCTGGTCAGTGCTGTTGGCCTCTGTCTTCTCTCTCGTTTGG 2930
           || |||||||||||| |||| |||||||||| ||||||||||||||||| ||  ||||
           601 ACTCCAACCATTGTTATTCTTTGGTCAGTGCTTTTGGCCTCTGTCTTCTCACTTATTTGG 660

      2931 GTCAAGAT 2938
           || |||||
           661 GTTAAGAT 668


arch completed: October 30, 2003, 16:25:42
b time : 6353.97 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 30, 2003, 16:71:59 ; Search Time 11509.5 Seconds
(without alignments)
11487.872 Million cell updates/sec

Title: US-09-980-043-1
Perfect score: 3232
Sequence: 1 gtcgacccacgcgtccgtct.....aaaaaaaaaaaaaaaaaaa 3232

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2888711 seqs, 20454813386 residues

Word size : 0

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3232	100.0	3232	8	AF072131	AF072131 Populus t
2	172	5.3	1010	8	AF197911	AF197911 Populus t
3	44	1.4	3145	8	AF323039	AF323039 Zinnia el
4	35	1.1	6886	8	AF413210	AF413210 Gossypium
c 5	32	1.0	130300	8	AP005824	AP005824 Oryza sat
6	32	1.0	146543	2	AP003945	AP003945 Oryza sat
c 7	32	1.0	153873	2	AL139160	AL139160 Homo sapi
8	32	1.0	208508	2	AC117774	AC117774 Mus muscu
9	31	1.0	1415	9	BC012630	BC012630 Homo sapi
10	31	1.0	3002	9	BC039244	BC039244 Homo sapi
11	31	1.0	109155	8	ATT20K12	AL137898 Arabidops
12	31	1.0	110000	2	PFMAL13_00	AL844509 Plasmodiu
13	31	1.0	111882	3	AC115612	AC115612 Dictyoste
14	31	1.0	153610	9	AC012443	AC012443 Homo sapi
15	31	1.0	160594	2	BX469920	BX469920 Danio rer
16	31	1.0	185011	2	BX511152	BX511152 Danio rer
c 17	30	0.9	81	6	AR119865	AR119865 Sequence
c 18	30	0.9	127	6	AR119867	AR119867 Sequence
c 19	30	0.9	176	6	AR119871	AR119871 Sequence
c 20	30	0.9	183	6	AR119868	AR119868 Sequence
c 21	30	0.9	200	6	AR119872	AR119872 Sequence
22	30	0.9	221	6	BD117641	BD117641 EST and e
c 23	30	0.9	238	14	D85022	D85022 Hepatitis C
c 24	30	0.9	309	6	AX407805	AX407805 Sequence
25	30	0.9	1159	8	AY035121	AY035121 Arabidops
26	30	0.9	1181	3	AK115493	AK115493 Ciona int
27	30	0.9	1250	8	AB012103	AB012103 Triticum
28	30	0.9	1944	9	HSM800162	AL049377 Homo sapi
29	30	0.9	3292	9	BC045542	BC045542 Homo sapi
30	30	0.9	4023	3	AY069707	AY069707 Drosophil
c 31	30	0.9	9518	6	AX100563	AX100563 Sequence
c 32	30	0.9	9595	6	AR119832	AR119832 Sequence
c 33	30	0.9	9595	14	AF054247	AF054247 Hepatitis
c 34	30	0.9	9595	14	AF054248	AF054248 Hepatitis
c 35	30	0.9	9596	14	AF054249	AF054249 Hepatitis
c 36	30	0.9	9599	6	AR119831	AR119831 Sequence
c 37	30	0.9	9599	6	AR119833	AR119833 Sequence
c 38	30	0.9	9599	14	AF011751	AF011751 Hepatitis
c 39	30	0.9	9599	14	AF011752	AF011752 Hepatitis
c 40	30	0.9	9599	14	AF011753	AF011753 Hepatitis
c 41	30	0.9	9611	6	AX057088	AX057088 Sequence
c 42	30	0.9	9611	6	AX057090	AX057090 Sequence
c 43	30	0.9	9611	6	AX057092	AX057092 Sequence
c 44	30	0.9	9611	6	AX057094	AX057094 Sequence
c 45	30	0.9	9611	14	AF177037	AF177037 Hepatitis

ALIGNMENTS

RESULT 1
AF072131
LOCUS AF072131 3232 bp mRNA linear PLN 26-APR-2001
DEFINITION Populus tremuloides secondary xylem cellulose synthase (celA) mRNA, complete cds.
ACCESSION AF072131
VERSION AF072131.1 GI:4115904
KEYWORDS
SOURCE Populus tremuloides (quaking aspen)
ORGANISM Populus tremuloides
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids I; Malpighiales; Salicaceae; Populus.
REFERENCE 1 (bases 1 to 3232)
AUTHORS Wu,L., Joshi,C.P. and Chiang,V.L.

```
TITLE     A xylem-specific cellulose synthase gene from aspen (Populus
          tremuloides) is responsive to mechanical stress
JOURNAL   Plant J. 22 (6), 495-502 (2000)
MEDLINE   20344632
 PUBMED   10886769
FERENCE   2  (bases 1 to 3232)
AUTHORS   Wu,L., Joshi,C.P. and Chiang,V.L.
TITLE     Direct Submission
JOURNAL   Submitted (14-JUN-1998) IWR/Forestry, Michigan Tech University,
          1400, Townsend Drive, Houghton, MI 49931, USA
FERENCE   3  (bases 1 to 3232)
AUTHORS   Wu,L., Joshi,C.P. and Chiang,V.L.
TITLE     Direct Submission
JOURNAL   Submitted (07-JAN-1999) IWR/Forestry, Michigan Tech University,
          1400, Townsend Drive, Houghton, MI 49931, USA
REMARK    Sequence update by submitter
MMENT     On Jan 8, 1999 this sequence version replaced gi:3244277.
ATURES             Location/Qualifiers
   source          1..3232
                   /organism="Populus tremuloides"
                   /mol_type="mRNA"
                   /db_xref="taxon:3693"
                   /tissue_type="secondary xylem"
   gene            1..3232
                   /gene="celA"
   CDS             69..3005
                   /gene="celA"
                   /codon_start=1
                   /product="secondary xylem cellulose synthase"
                   /protein_id="AAD03417.1"
                   /db_xref="GI:4115905"
                   /translation="MMESGAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEI
                   NEGRKVCLRCGSPYDENLLDDVEKKGSGNQSTMASHLNDSQDVGIHARHISSVSTVDS
                   EMNDEYGNPIWKNRVKSCKDKENKKKKRSPKAETEPAQVPTEQQMEEKPSAEASEPLS
                   IVYPIPRNKLTPYRAVIIMRLVILGLFFHFRITNPVDSAFGLWLTSVICEIWFAFSWV
                   LDQFPKWNPVNRETYIERLSARYEREGEPSQLAGVDFFVSTVDPLKEPPLITANTVLS
                   ILAVDYPVDKVSCYVSDDGAAMLSFESLVETAEFARKWVPFCKKFSIEPRAPEFYFSQ
                   KIDYLKDKVQPSFVKERRAMKRDYEEYKVRVNALVAKAQKTPEEGWTMQDGTPWPGNN
                   TRDHPGHDSGLPWEILGARDIEGNELPRLVYVSREKRPGYQHHKKAGAENALVRVSAV
                   LTNAPYILNVDCDHYVNNSKAVREAMCILMDPQVGRDVCYVQFPQRFDGIDKSDRYAN
                   RNVVFFDVNMKGLDGIQGPVYVGTGCVFNRQALYGYGPPSMPSLRKRKDSSSCFSCCC
                   PSKKKPAQDPAEVYRDAKREDLNAAIFNLTEIDNYDEHERSMLISQLSFEKTFGLSSV
                   FIESTLMENGGVPESANSPPFIKEAIQVIGCGYEEKTEWGKQIGWIYGSVTEDILSGF
                   KMHCRGWRSIYCMPVRPAFKGSAPINLSDRLHQVLRWALGSVEIFFSRHCFLWYGFGG
                   GRLKWLQRLAYINTIVYPFTSLPLIAYCTIPAVCLLTGKFIIPTLSNLASMLFLGLFI
                   SIIVTAVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQGFLKMLAGIDTNFTVTA
                   KAAEDAEFGELYMVKWTTLLIPPTTLLIINMSGCAGFSDALNKGYEAWGPLFGKVFFA
                   FWVILHLYPFLKGLMGRQNLTPTIVVLWSVLLASVFSLVWVKINPFVNKVDNTLVAET
                   CISIDC"
SE COUNT      895 a    668 c    772 g    897 t
IGIN

Query Match             100.0%;  Score 3232;  DB 8;  Length 3232;
Best Local Similarity   100.0%;  Pred. No. 0;
Matches 3232;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

          1 GTCGACCCACGCGTCCGTCTTGAAAGAATATGAAGTTGTAAAGAGCTGGTAAAGTGGTAA 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
          1 GTCGACCCACGCGTCCGTCTTGAAAGAATATGAAGTTGTAAAGAGCTGGTAAAGTGGTAA 60

         61 TAAGCAAGATGATGGAATCTGGGGCTCCTATATGCCATACCTGTGGTGAACAGGTGGGGC 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
         61 TAAGCAAGATGATGGAATCTGGGGCTCCTATATGCCATACCTGTGGTGAACAGGTGGGGC 120

        121 ATGATGCAAATGGGGAGCTATTTGTGGCTTGCCATGAGTGTAGCTATCCCATGTGCAAGT 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
        121 ATGATGCAAATGGGGAGCTATTTGTGGCTTGCCATGAGTGTAGCTATCCCATGTGCAAGT 180

        181 CTTGTTTCGAGTTTGAAATCAATGAGGGCCGGAAAGTTTGCTTGCGGTGTGGCTCGCCAT 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
        181 CTTGTTTCGAGTTTGAAATCAATGAGGGCCGGAAAGTTTGCTTGCGGTGTGGCTCGCCAT 240

        241 ATGATGAGAACTTGCTGGATGATGTAGAAAAGAAGGGGTCTGGCAATCAATCCACAATGG 300
```

```
Db        241 ATGATGAGAACTTGCTGGATGATGTAGAAAAGAAGGGGTCTGGCAATCAATCCACAATGG 300

Qy        301 CACTTTACAGTCAA                                                 360
              |||  |||  |||                                                  |||
Db        301 CATCTTACAGTCAA                                                 360

Qy        361 CCACTGTGGATAGT                                                 420
              |||  |||  |||                                                  |||
Db        361 CCACTGTGGATAGT                                                 420

Qy        421 AGAGCTTAAAGACTAAAAAAAAAAAAAAAAAAAAAAAAAGAAGTCCTAAGGCTGAAACTGAAC 480
              |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db        421 AGAGCTTAAAGACTAAAAAAAAAAAAAAAAAAAAAAAAAGAAGTCCTAAGGCTGAAACTGAAC 480

Qy        481 CAGCTCAAGTTCGCTAAAAAAAAAAAAAAAAAAAAAAAAAACGTCTGCAGAGGCTCGGAGC 540
              |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db        481 CAGCTCAAGTTCGCTAAAAAAAAAAAAAAAAAAAAAAAAAACGTCTGCAGAGGCTCGGAGC 540

Qy        541 CGCTTTCAATCTCTCTAATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCA 600
              |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db        541 CGCTTTCAATCTCTCTAATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCA 600

Qy        601 TTATGCGACTGGCTCTCTCGTTCTTCTTTCTCTACTTCAGAATAACAAATCCCTCTCGATA 660
              |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db        601 TTATGCGACTGGCTCTCTCGTTCTTCTTTCTCTACTTCAGAATAACAAATCCCTCTCGATA 660

Qy        661 GTGCCTTTGGCCTCTCTCCTTACTTTCTCTATATGTGAGATCTGGTTTGCATTTCTTTGGG 720
              |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db        661 GTGCCTTTGGCCTCTCTCCTTACTTTCTCTATATGTGAGATCTGGTTTGCATTTCTTTGGG 720

Qy        721 TGTTGGATTAGTTTCTCAAGTCGAATCCTCTTAATAGAGAAACGTATATCGAAAGGCTGT 780
              |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db        721 TGTTGGATTAGTTTCTCAAGTCGAATCCTCTTAATAGAGAAACGTATATCGAAAGGCTGT 780

Qy        781 CGGCAACGTATGAAAGAGAGCGTGACTCTTTCTACTTGCTGGTGTGGATTTTTTCTGTGA 840
              |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db        781 CGGCAACGTATGAAAGAGAGCGTGACTCTTTCTACTTGCTGGTGTGGATTTTTTCTGTGA 840

Qy        841 GTACTGTTGATCCCTTGAAGGAACCGGTCATTTACTGCCAATACAGTCCTTTCCATCC 900
              |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db        841 GTACTGTTGATCCCTTGAAGGAACCGGTCATTTACTGCCAATACAGTCCTTTCCATCC 900

Qy        901 TTGCTGTGGACTATCTGTGGTGATAAAGTCTCTTGCTACGTGTCTGATGATGGTGCAGCTA 960
              |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db        901 TTGCTGTGGACTATCTGTGGTGATAAAGTCTCTTGCTACGTGTCTGATGATGGTGCAGCTA 960

Qy        961 TGCTTTCATTTGAATTCTTGTAGAAACACTCGAGTTTGCAAGGAAGTGGGTTCCGTTCT 1020
              |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db        961 TGCTTTCATTTGAATTCTTGTAGAAACACTCGAGTTTGCAAGGAAGTGGGTTCCGTTCT 1020

Qy       1021 GCAAAAAAATTGTCAATTGAAGCAAGAGCACTCGAGTTTACTTCTCACAGAAAATTGATT 1080
              |||||||||||||  |||  |||  |||  |||||||||||||||||||||||||||||||
Db       1021 GCAAAAAAATTGTCAATTGAAGCAAGAGCACTCGAGTTTACTTCTCACAGAAAATTGATT 1080

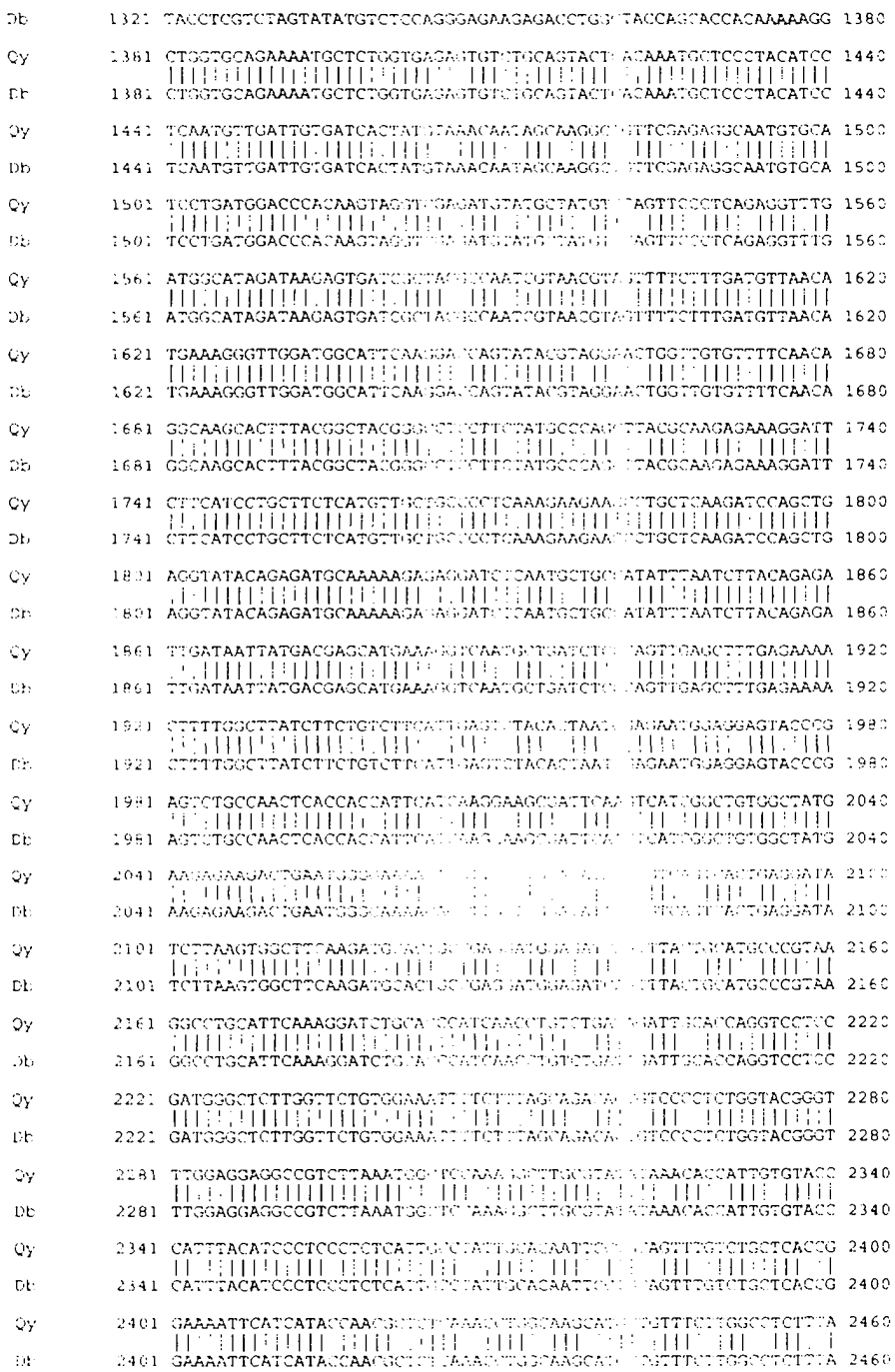
Qy       1081 ACTTGAAAGACAAGTTCAACTTCTTCTTTCTCTAAAGAACGTAGAGCAATGAAAAGGGATT 1140
              |||  |||  |||  |||  |||  |||  |||  |||||||||||||||||||||||||
Db       1081 ACTTGAAAGACAAGTTCAACTTCTTCTTTCTCTAAAGAACGTAGAGCAATGAAAAGGGATT 1140

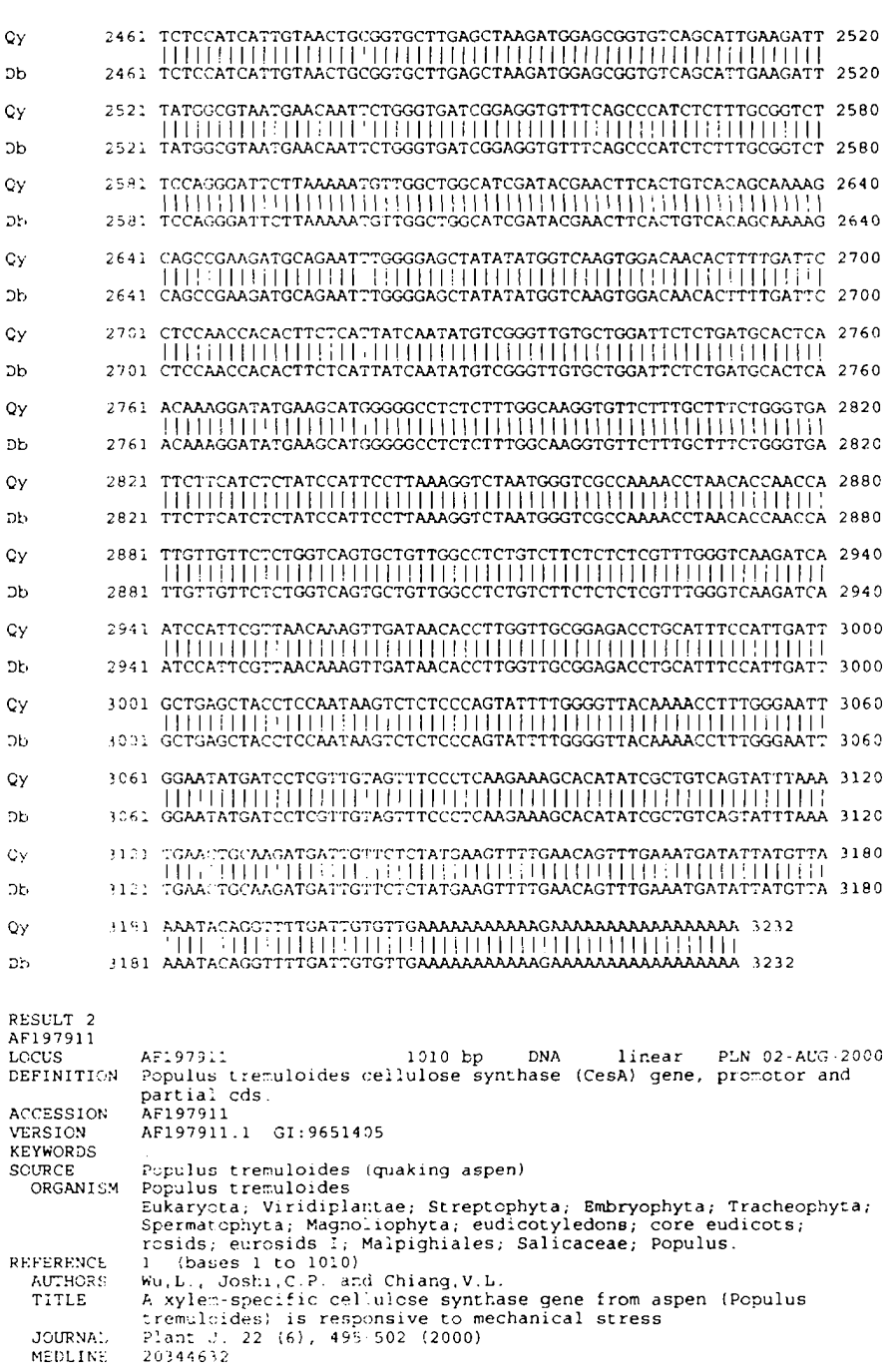
Qy       1141 ATGAAGAGTACAAATCCGAGTTAATTGCTTTAGCAAAGGCTCAGAAAACACCTGAAG 1200
              |||||||||||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db       1141 ATGAAGAGTACAAATCCGAGTTAATTGCTTTAGCAAAGGCTCAGAAAACACCTGAAG 1200

Qy       1201 AAGGATGGACTATCAAGATGGAACACCTTGGCCTGGGAATAACACACGTGATCACCCTG 1260
              |||||||  |||  |||  |||||||||||  |||  |||  |||||||||||  |||||||
Db       1201 AAGGATGGACTATCAAGATGGAACACCTTGGCCTGGGAATAACACACGTGATCACCCTG 1260

Qy       1261 GGCATGATTCAGGTTTTCCTTGGGAAATATTGGGAGCTCGTGACATTGAAGGAAATGAAC 1320
              |||||||||||||  |||  |||  |||  |||  |||  |||||||||||  |||||||
Db       1261 GGCATGATTCAGGTTTTCCTTGGGAAATATTGGGAGCTCGTGACATTGAAGGAAATGAAC 1320

Qy       1321 TACCTCGTCTAGTATCATGTCTCCAGGGAGAAGAGACCTGGCTACCAGCACCACAAAAAGG 1380
              |||||||||||||  |||||||||||  |||||||||||||||  |||||||  |||||
Db       1321 TACCTCGTCTAGTATCATGTCTCCAGGGAGAAGAGACCTGGCTACCAGCACCACAAAAAGG 1380
```





RESULT 2
AF197911
LOCUS AF197911 1010 bp DNA linear PLN 02-AUG-2000
DEFINITION Populus tremuloides cellulose synthase (CesA) gene, promotor and partial cds.
ACCESSION AF197911
VERSION AF197911.1 GI:9651405
KEYWORDS .
SOURCE Populus tremuloides (quaking aspen)
ORGANISM Populus tremuloides
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids I; Malpighiales; Salicaceae; Populus.
REFERENCE 1 (bases 1 to 1010)
AUTHORS Wu,L., Joshi,C.P. and Chiang,V.L.
TITLE A xylem-specific cellulose synthase gene from aspen (Populus tremuloides) is responsive to mechanical stress
JOURNAL Plant J. 22 (6), 495-502 (2000)
MEDLINE 20944632

```
 PUBMED   10886769
FERENCE   2  (bases 1 to 1010)
AUTHORS   Wu,L., Joshi,C.P. and Chiang,V.L.
TITLE     Direct Submission
JOURNAL   Submitted (22-OCT-1999) Forestry, Michigan Tech University, 1400,
          Townsend Drive, Houghton, MI 49931, USA
ATURES             Location/Qualifiers
   source          1. .1010
                   /organism="Populus tremuloides"
                   /mol_type="genomic DNA"
                   /db_xref="taxon:3693"
                   /tissue_type="xylem"
   gene            <1. .>1010
                   /gene="CesA"
   misc_feature    1. .840
                   /gene="CesA"
                   /note="contains promotor and 5'UTR"
   mRNA            <841. .>1010
                   /gene="CesA"
                   /product="cellulose synthase"
   CDS             841. .>1010
                   /gene="CesA"
                   /codon_start=1
                   /product="cellulose synthase"
                   /protein_id="AAF91155.1"
                   /db_xref="GI:9651406"
                   /translation="MMESGAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEI
                   KEGRKVCLRCGS"
SE COUNT      289 a    212 c    216 g    293 t
IGIN

Query Match             5.3%;  Score 172;  DB 8;  Length 1010;
Best Local Similarity   100.0%;  Pred. No. 1.9e-87;
Matches  172;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

         31 TGAAGTTGTAAAGAGCTGGTAAAGTGGTAATAAGCAAGATGATGGAATCTGGGGCTCCTA 90
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
        803 TGAAGTTGTAAAGAGCTGGTAAAGTGGTAATAAGCAAGATGATGGAATCTGGGGCTCCTA 862

         91 TATGCCATACCTGTGGTGAACAGGTGGGGCATGATGCAAATGGGGAGCTATTTGTGGCTT 150
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
        863 TATGCCATACCTGTGGTGAACAGGTGGGGCATGATGCAAATGGGGAGCTATTTGTGGCTT 922

        151 GCCATGAGTGTAGCTATCCCATGTGCAAGTCTTGTTTCGAGTTTGAAATCAA 202
            |||||||||||||||||||||||||||||||||||||||||||||||||||
        923 GCCATGAGTGTAGCTATCCCATGTGCAAGTCTTGTTTCGAGTTTGAAATCAA 974


SULT 3
323039
CUS       AF323039               3145 bp    mRNA    linear   PLN 25-FEB-2002
FINITION  Zinnia elegans cellulose synthase CesA-1 mRNA, complete cds.
CESSION   AF323039
RSION     AF323039.2  GI:18875453
YWORDS    .
URCE      Zinnia elegans
ORGANISM  Zinnia elegans
          Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
          Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
          Asteridae; campanulids; Asterales; Asteraceae; Astercideae;
          Heliantheae; Zinnia.
FERENCE   1  (bases 1 to 3145)
AUTHORS   Haigler,C.H., Ivanova-Datcheva,M., Hogan,P.S., Salnikov,V.V.,
          Hwang,S., Martin,K. and Delmer,D.P.
TITLE     Carbon partitioning to cellulose synthesis
JOURNAL   Plant Mol. Biol. 47 (1-2), 29-51 (2001)
MEDLINE   21438255
 PUBMED   11554477
FERENCE   2  (bases 1 to 3145)
AUTHORS   Hwang,S. and Haigler,C.H.
TITLE     Direct Submission
JOURNAL   Submitted (22-NOV-2000) Biological Sciences, Texas Tech University,
          Box 43131, Lubbock, TX 79409, USA
```

```
REFERENCE   3  (bases 1 to [illegible])
  AUTHORS   Hwang,S. and Ha[illegible],C.H.
  TITLE     Direct Submiss[illegible]
  JOURNAL   Submitted (25 FE[illegible] 2002) Biological Sciences, Texas Tech University,
            Box 43131, Lubb[illegible]k, TX 7940[illegible], USA
  REMARK    Sequence update by submitter
COMMENT     On Feb 25, 2002 this sequence version replaced gi:13021930.
FEATURES             Locati[illegible] Qualifiers
     source          1. .[illegible]
                     /organ[illegible]="Zinnia elegans"
                     /mol_ty[illegible]="mRNA"
                     /db_xre[illegible]="taxon:34745"
     CDS             69. .[illegible]
                     /codon[illegible]
                     /produ[illegible]="cellulose synthase CesA-1"
                     /prote[illegible]id "AAK11588.1"
                     /db_xref="GI:18875454"
                     /transl[illegible]="MMESPETHICNICGENIGFGPDGNGVFVGCHECNFPICKSCLEYE
                     INPGNTA[illegible]SIEKMPKEPVTRNTMAAHLSNSQDAGLHARNVSTVSTV
                     QTPM[illegible]SIIKAMPVPAAS[illegible]KLKEFFGSDKVAKEVQVPEDQHTEEKQQSADPNAM
                     [illegible]  [illegible] ILPFHYSITNPVESSYSLWLTSVICEIWFA
                     LSWVID[illegible]EATPLAMPAFYLSHIIPPYFRELEFSELAAVDLFFVSTVLPLKEPFLITAN
                     TVIS[illegible]IPVFPSMVTLIL[illegible]AFMLTPFSLVFTAEFARKWVPFCKKFSIEPRAPEF
                     YFSQR[illegible]KEKVLFSPVFELFAMSPLYEEYKVRVNALVAKAQKTPEEGWTMQDGTPW
                     PGNNPR[illegible]GMIQVFLGHSGAHDTFGNELPRLVYVSREKRPGYQHHKKAGAENALVRV
                     SAVLTNAPYILNLDCDHYVKNSKAVREAMCFMMDPQVGRDVCYIQFPQRFDGIDRSDR
                     YANRNTVFFDVNMKGLDGIQGPVYVGTGCVFYRQALYGYGPPCSLPTLPSPSSSSSCCC
                     CGPKKPKKDGLEEFKRDARPDDLNAAIFNLKEIESYDDYERSLLISQMSFEKTFGMSSV
                     FIESTLMENGGLAESANPATMINEAIHVISCGYEEKTAWGKEIGWIYGSVTEDILTGF
                     KMHCRGWRSIYCMPVRPAFKGSAPINLSDRLHQVLRWALGSVEIFLSRHCPLWYGWGG
                     GRLKLLQRIAYINTIVYPFTSIPLVAYCTLPAICLLTGKFIIPTLSNIAAVWFIGLFL
                     SIITTSVLEIRWSGVSIEELWRNEQFWVIGGVSAHLFAVFQGFLKMLAGVDTNFTVTA
                     KAADDQEF HELYMIKWTTVLIPPTTLLVLNLVGVVAGFSDALNKGYEAWGPLFGKVFF
                     AFWVIL[illegible]PFLKGLMGRQNRTPTIVILWSVLLASVFSLVWVKIDPFVSKQDSNITQS
                     CIAID[illegible]
BASE COUNT      873 a    582 c    746 g    944 t
ORIGIN

  Query Match             1.4%;  Score 44;  DB 8;  Length 3145;
  Best Local Similarity   1[illegible] 0%;  Pred. No. [illegible]e-13;
  Matches   44;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1194 CCTGAAGAAGGATT ACTATGCAAGATGGAA ACCTTGGCCTGG 1237
              ||||||||||||||  ||| ||| ||| |||||||||||||
Db       1206 CCTGAAGAAGGATT ACTATGCAAGATGGAA ACCTTGGCCTGG 1249


RESULT 4
AF413210
LOCUS       AF413210               6986 bp    DNA     linear   PLN 27-DEC-2001
DEFINITION  Gossypium hirsutum cellulose synthase A4 (CesA4) gene, complete
            cds.
ACCESSION   AF413210
VERSION     AF413210.1  GI:[illegible]226293
KEYWORDS    .
SOURCE      Gossypium hirsutum (upland cotton)
  ORGANISM  Gossypium hirsutum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium.
REFERENCE   1  (bases 1 to 6986)
  AUTHORS   Kim,H.J. and Triplett,B.A.
  TITLE     Cotton fiber growth in planta and in vitro. Models for plant cell
            elongation and cell wall biogenesis
  JOURNAL   Plant Physiol. [illegible] (4), 1361-1366 (2001)
  MEDLINE   21608363
   PUBMED   11743074
REFERENCE   2  (bases 1 to 6986)
  AUTHORS   Kim,H.J., Williams,M.Y. and Triplett,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-AUG-2001) Fiber Bioscience, USDA-ARS, SRRC, 1100
            Robert E. Lee Blvd, New Orleans, LA 70124, USA
FEATURES             Location/Qualifiers
```

```
     source          1. .6886
                     /organism="Gossypium hirsutum"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3635"
     gene            <2685. .>6766
                     /gene="CesA4"
     mRNA            join(<2685. .2862,3033. .3106,3193. .3244,3338. .3482,
                     3613. .3879,3992. .4337,4426. .4663,4778. .4930,5085. .5328,
                     5420. .5619,5732. .6085,6179. .6766)
                     /gene="CesA4"
                     /product="cellulose synthase A4"
     CDS             join(2685. .2862,3033. .3106,3193. .3244,3338. .3482,
                     3613. .3879,3992. .4337,4426. .4663,4778. .4930,5085. .5328,
                     5420. .5619,5732. .6085,6179. .6766)
                     /gene="CesA4"
                     /codon_start=1
                     /product="cellulose synthase A4"
                     /protein_id="AAL87718.1"
                     /db_xref="GI:17226294"
                     /translation="MMESGVPVCHTCGEHVGLNVNGEPFVACHECNFPICKSCFEYDL
                     KEGRKACLRCGSPYDENLLDDVEKTTGDQSTMAAHLSKSQDVGIHARHISSVSTLDSE
                     MTGDNGNPIWKNRVESWKEKKNKKKKPATTKVEFEAEIPPEQQMEDKPAPDASQPLST
                     IIPIPKSRLAPYRTVIIMRLIILGLFFHYRVTHPVDSAFGLWLTSVICEIWFAFSWVL
                     DQFPKWYPVNRETYIDRLSARYDREGEPNELAAVDFFVSTVDPLKEPPLITANTVLSI
                     LALDYPVDKVSCYISDDGAAMLTFESLVETADFARKWVPFCKKFSIEPRAPEFYFSQK
                     IVYLKDKVQPSFVKERRAMKRDYEEYKIRINALVAKAQKTPEEGWTMQDGTPWPGNNP
                     RDHPGMIQVFLGYSGAHDIEGNELPRLVYVSREKRPGYQHHKQAGADNALVRVSAALT
                     NAPFILNLDCDHYVNNSKAVREAMCFLMDPQGGKDVCYVQFPQRFDGIDRSERYAKRN
                     TVFFDVNMKGRDGSQGPVYVGTGCVFNRQALYGYDPPSMPSFPKSSSSSCSCCCPGKK
                     EPKEPTELYRDAKREEDAAIFNLREIDNYDEYERSMLISQTSFEKTFGLSSVFIEST
                     LMENGGVAESANPSTLIKEAIHVISCGYEEKTAWGKEIGWIYGSVTEDILTGFKMHCR
                     GWRSIYCMPLRPAFKGSAPINLSDRLHQVLRWALGSLEIFLSRHCPLWYGFGGGRLKW
                     LQRLAYINTSVYPFTSIPLIAYCSLPAICLLTGKFIIPTLSNLASVLYLGLFLSIIVT
                     AVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQGFLKVLAGIDTNFTVTAKAADD
                     ADFGELYIVKWTTLLIPPTTLLIVNMVGVVAGFSLALNKGYEAWGPLFGKVFFSFWVI
                     LHLYPFLKGLMGRQNRTPTIVVLWSVLLASVFSLVWVRINPFVSTADSTTVSRANSSI
                     DC"
BASE COUNT     2067 a    1237 c    1277 g    2304 t       1 others
ORIGIN

  Query Match             1.1%;  Score 35;  DB 8;  Length 6886;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-07;
  Matches   35;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1176 GCAAAGGCTCAGAAAACACCTGAAGAAGTATGGAT 1211
             ||||||||||||||||||||||||||||||||||||
Db       4468 GCAAAGGCTCAGAAAACACCTGAAGAAGTATGGAT 4503


RESULT 5
AP005824/c
LOCUS       AP005824               130300 bp    DNA     linear   PLN 05-JUN-2003
DEFINITION  Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 7,
            PAC clone:P0669H03, complete sequence.
ACCESSION   AP005824
VERSION     AP005824.2  GI:31414615
KEYWORDS    HTG.
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1
  AUTHORS   Sasaki,T., Matsumoto,T. and Katayose,Y.
  TITLE     Oryza sativa nipponbare(GA3) genomic DNA, chromosome 7, PAC
            clone:P0669H03
  JOURNAL   Published Only in Database (2002)
REFERENCE   2  (bases 1 to 130300)
  AUTHORS   Sasaki,T., Matsumoto,T. and Katayose,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-OCT-2002) Takuji Sasaki, National Institute of
            Agrobiological Sciences, Rice Genome Research Program, Kannondai
            2-1-2, Tsukuba, Ibaraki 305-8602, Japan
            (E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
            Tel:81-298-38-7441, Fax:81-298-38-7468)
  COMMENT   On Jun 4, 2003 this sequence version replaced gi:23821522.
            The orientation of the sequence is from SP6 to T7 of the PAC clone.
  FEATURES             Location/Qualifiers
       source          1. .130300
                       /organism="Oryza sativa (japonica cultivar-group)"
                       /mol_type="genomic DNA"
                       /cultivar="Nipponbare"
                       /db_xref="taxon:39947"
                       /chromosome="7"
                       /clone="P0669H03"
  BASE COUNT    37609 a   28033 c   28472 g   36186 t
  ORIGIN

  Query Match             1.0%;  Score 32;  DB 8;  Length 130300;
  Best Local Similarity   100.0%;  Pred. No. 6.9e-06;
  Matches   32;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1533 TGCTATGTGCAGTTCCCTCAGAGGTTTGATGG 1564
              ||||||||||||||||| ||||||||||||
Db     103857 TGCTATGTGCAGTTCCCTCAGAGGTTTGATGG 103826


RESULT 6
AP003945
LOCUS       AP003945               146543 bp    DNA     linear   HTG 21-MAR-2002
DEFINITION  Oryza sativa (japonica cultivar-group) chromosome 7 clone
            OJ1134_A07, *** SEQUENCING IN PROGRESS ***.
ACCESSION   AP003945
VERSION     AP003945.1  GI:15021915
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Oryza sativa nipponbare(GA3) genomic DNA, chromosome 7, BAC
            clone:OJ1134_A07
  JOURNAL   Published Only in Database (2001)
REFERENCE   2  (bases 1 to 146543)
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JUL-2001) Takuji Sasaki, National Institute of
            Agrobiological Resources, Rice Genome Research Program; Kannondai
            2-1-2, Tsukuba, Ibaraki 305-8602, Japan
            (E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
            Tel:81-298-38-7441, Fax:81-298-38-7468)
  COMMENT   The nucleotide sequence of this BAC clone was generated by
            combining Monsanto and RGP-Japan sequencing data.
            NOTE: It currently consists of 1 contigs. Gaps between the contigs
            are represented as runs of N. The order of the pieces is believed
            to be correct as given, however the sizes of the gaps between them
            are based on estimates that have provided by the submitter. This
            sequence will be replaced by the finished sequence as soon as it is
            available and the accession number will be preserved.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
  FEATURES             Location/Qualifiers
       source          1. .146543
                       /organism="Oryza sativa (japonica cultivar-group)"
                       /mol_type="genomic DNA"
                       /cultivar="Nipponbare"
                       /db_xref="taxon:39947"
                       /chromosome="7"
                       /clone="OJ1134_A07"
  BASE COUNT    41566 a   31160 c   31453 g   42148 t     216 others
  ORIGIN
```

```
uery Match              1.0%;  Score 32;  DB 2;  Length 146543;
est Local Similarity   100.0%;  Pred. No. 6.9e-06;
atches   32;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

      1533 TGCTATGTGCAGTTCCCTCAGAGGTTTGATGG 1564
           ||||||||||||||||||||||||||||||||
     80563 TGCTATGTGCAGTTCCCTCAGAGGTTTGATGG 80594


ULT 7
39160/c
US       AL139160              153873 bp    DNA     linear   HTG 15-MAY-2002
INITION  Homo sapiens chromosome 1 clone RP1-93K19, *** SEQUENCING IN
         PROGRESS ***, 25 unordered pieces.
ESSION   AL139160
SION     AL139160.3  GI:9796348
WORDS    HTG; HTGS_PHASE1; HTGS_CANCELLED.
RCE      Homo sapiens (human)
RGANISM  Homo sapiens
         Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
         Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
ERENCE   1  (bases 1 to 153873)
UTHORS   McLay,K.
ITLE     Direct Submission
OURNAL   Submitted (14-MAY-2002) Wellcome Trust Sanger Institute, Hinxton,
         Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
         humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
MENT     On Aug 11, 2000 this sequence version replaced gi:9212199.
          ------------- Genome Center
         Center: Wellcome Trust Sanger Institute
         Center code: SC
         Web site: http://www.sanger.ac.uk
         Contact: humquery@sanger.ac.uk
         -------------- Project Information
         Center project name: dJ93K19
         - ------------ Summary Statistics
         Assembly program: XGAP4; version 4.5
         Sequencing vector: plasmid; L08752; 100% of reads
         Chemistry: Dye-terminator Big Dye; 100% of reads
         Consensus quality: 138597 bases at least Q40
         Consensus quality: 143343 bases at least Q30
         Consensus quality: 146407 bases at least Q20
         Insert size: 151473; sum-of-contigs
         Quality coverage: 3.36x in Q20 bases; sum-of-contigs
         ---------------
         * NOTE: This is a 'working draft' sequence. It currently
         * consists of 25 contigs. The true order of the pieces
         * is not known and their order in this sequence record is
         * arbitrary. Gaps between the contigs are represented as
         * runs of N, but the exact sizes of the gaps are unknown.
         * This record will be updated with the finished sequence
         * as soon as it is available and the accession number will
         * be preserved.
         *        1     2675: contig of 2675 bp in length
         *     2676     2775: gap of 100 bp
         *     2776    10981: contig of 8206 bp in length
         *    10982    11081: gap of 100 bp
         *    11082    19149: contig of 8068 bp in length
         *    19150    19249: gap of 100 bp
         *    19250    22690: contig of 3441 bp in length
         *    22691    22790: gap of 100 bp
         *    22791    26878: contig of 4088 bp in length
         *    26879    26978: gap of 100 bp
         *    26979    37337: contig of 10359 bp in length
         *    37338    37437: gap of 100 bp
         *    37438    41551: contig of 4114 bp in length
         *    41552    41651: gap of 100 bp
         *    41652    44476: contig of 2825 bp in length
         *    44477    44576: gap of 100 bp
         *    44577    46980: contig of 2404 bp in length
         *    46981    47080: gap of 100 bp
         *    47081    57594: contig of 10514 bp in length
         *    57595    57694: gap of 100 bp
         *    57695    59862: contig of 2168 bp in length
         *    59863    59962: gap of 100 bp
         *    59963    63882: contig of 3920 bp in length
         *    63883    63982: gap of 100 bp
         *    63983    80635: contig of 16653 bp in length
         *    80636    80735: gap of 100 bp
         *    80736    90675: contig of 9940 bp in length
         *    90676    90875: gap of 100 bp
         *    90876    93940: contig of 3065 bp in length
         *    93941    94040: gap of 100 bp
         *    94041    96284: contig of 2244 bp in length
         *    96285    96384: gap of 100 bp
         *    96385    99200: contig of 2816 bp in length
         *    99201    99300: gap of 100 bp
         *    99301   104274: contig of 4974 bp in length
         *   104275   104374: gap of 100 bp
         *   104375   110768: contig of 6394 bp in length
         *   110769   110868: gap of 100 bp
         *   110869   113245: contig of 2377 bp in length
         *   113246   113345: gap of 100 bp
         *   113346   116491: contig of 3146 bp in length
         *   116492   116591: gap of 100 bp
         *   116592   131479: contig of 14888 bp in length
         *   131480   131579: gap of 100 bp
         *   131580   140766: contig of 9187 bp in length
         *   140767   140866: gap of 100 bp
         *   140867   146823: contig of 5957 bp in length
         *   146824   146923: gap of 100 bp
         *   146924   153873: contig of 6950 bp in length.
FEATURES             Location/Qualifiers
     source          1..153873
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP1-93K19"
                     /clone_lib="RPCI-3"
     misc_feature    1..2675
                     /note="assembly fragment:2175
                     fragment_chain:1
                     clone_end:SP6
                     vector_side:left"
     misc_feature    2776..10981
                     /note="assembly fragment:8198
                     fragment_chain:1"
     misc_feature    11082..19149
                     /note="assembly fragment:8170
                     fragment_chain:1"
     misc_feature    19250..22690
                     /note="assembly_fragment:3416
                     fragment_chain:1"
     misc_feature    22791..26878
                     /note="assembly_fragment:4088
                     fragment_chain:1"
     misc_feature    26979..37337
                     /note="assembly_fragment:10359
                     fragment_chain:1"
     misc_feature    37438..41551
                     /note="assembly fragment:4114
                     fragment_chain:1"
     misc_feature    41652..44476
                     /note="assembly fragment:2825
                     fragment_chain:2"
     misc_feature    44577..46980
                     /note="assembly_fragment:2404
                     fragment_chain:2"
     misc_feature    47081..57594
                     /note="assembly_fragment:10514
                     fragment_chain:2"
     misc_feature    57695..59862
                     /note="assembly_fragment:2168
                     fragment_chain:3"
     misc_feature    59963..63882
```

```
                     /note="assembly_fragment:00647
                     fragment_chain:3"
     misc_feature    63983. .80635
                     /note="assembly_fragment:01175
                     fragment_chain:4"
     misc_feature    80736. .90775
                     /note="assembly_fragment:00630
                     fragment_chain:4"
     misc_feature    90876. .93940
                     /note="assembly_fragment:01333
                     fragment_chain:5"
     misc_feature    94041. .96283
                     /note="assembly_fragment:01135
                     fragment_chain:5"
     misc_feature    96384. .99199
                     /note="assembly_fragment:00089"
     misc_feature    99300. .104274
                     /note="assembly_fragment:00110"
     misc_feature    104375. .110797
                     /note="assembly_fragment:00283"
     misc_feature    110898. .113380
                     /note="assembly_fragment:00378"
     misc_feature    113481. .116491
                     /note="assembly_fragment:00409"
     misc_feature    116592. .131479
                     /note="assembly_fragment:00761"
     misc_feature    131580. .140766
                     /note="assembly_fragment:00453"
     misc_feature    140867. .146823
                     /note="assembly_fragment:00964"
     misc_feature    146924. .153873
                     /note="assembly_fragment:00375
                     clone_end:3'
                     vector_side:right"
BASE COUNT    41026 a   34466 c   34931 g   40955 t    24[illegible] others
ORIGIN

  Query Match             1.0%;  Score 32;  DB 2;  Length 153873;
  Best Local Similarity   100.0%;  Pred. No. 6.9e-06;
  Matches   32;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3221 TTGAAAAAAAAAAAGAAAAAAAAAAAAAAAAA 3252
             ||||||||||||||||||||||||||||||||
Db      12915 TTGAAAAAAAAAAAGAAAAAAAAAAAAAAAAA 12884


RESULT 6
AC117774
LOCUS       AC117774               [illegible] bp    DNA     linear   HTG 17-MAR-2003
DEFINITION  Mus musculus clone RP24 [illegible] WORKING DRAFT SEQUENCE, 4 unordered
            pieces.
ACCESSION   AC117774
VERSION     AC117774.4  GI:28975961
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 209508)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP24 [illegible]
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 208508)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Carvalho,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colampolo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Glinde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,E., Lamberson,R.,
            Landers,T., Lebovitz,J., Levine,R., Lindblad-Toh,K., Liu,G.,
```

```
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zaincun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-APR-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 208508)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zaincun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAR-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 17, 2003 this sequence version replaced gi:27923677.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www.seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L25804
                Center clone name: 534_P_19
            -------------- Summary Statistics
                Sequencing vector: Plasmid; n/a; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 207505 bases at least Q40
                Consensus quality: 207970 bases at least Q30
                Consensus quality: 208122 bases at least Q20
                Insert size: 200000; agarose-fp
                Insert size: 209208; sum-of-contigs
                Quality coverage: 10.2 in Q20 bases; agarose-fp
                Quality coverage: 9.8 in Q20 bases; sum-of-contigs
            --------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved
            *       1   124039  contig of 124039 bp in length
            *  124040   124139  gap of 100 bp
```

```
          •   124140   132419: contig of 8280 bp in length
          •   132420   132519: gap of 100 bp
          •   132520   157397: contig of 24878 bp in length
          •   157398   157497: gap of 100 bp
          •   157498   208508: contig of 51011 bp in length.
ATURES             Location/Qualifiers
   source          1. .208508
                   /organism="Mus musculus"
                   /mol_type="genomic DNA"
                   /db_xref="taxon:10090"
                   /clone="RP24-534P19"
                   /clone_lib="RPCI-24 Male Mouse BAC"
   misc_feature    1. .124039
                   /note="assembly_fragment
                   clone_end:SP6
                   vector_side:left"
   misc_feature    124140. .132419
                   /note="assembly_fragment"
   misc_feature    132520. .157397
                   /note="assembly_fragment"
   misc_feature    157498. .208508
                   /note="assembly_fragment"
SE COUNT    57297 a  45581 c  45722 g  59607 t    301 others
IGIN

Query Match             1.0%;  Score 32;  DB 2;  Length 208508;
Best Local Similarity   100.0%;  Pred. No. 7e-06;
Matches   32;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

       3201 TTGAAAAAAAAAAAAGAAAAAAAAAAAAAAAAA 3232
            ||||:|.||||||||||||||||||||||||||
      93868 TTGAAAAAAAAAAAAGAAAAAAAAAAAAAAAAA 93899


SULT 9
012630
CUS       BC012630               1415 bp    mRNA    linear   PRI 20 AUG-2001
FINITION  Homo sapiens, clone IMAGE:4282420, mRNA.
CESSION   BC012630
RSION     BC012630.1  GI:15215008
YWORDS    .
URCE      Homo sapiens (human)
ORGANISM  Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
FERENCE   1  (bases 1 to 1415)
AUTHORS   Strausberg,R.
TITLE     Direct Submission
JOURNAL   Submitted (15-AUG-2001) National Institutes of Health, Mammalian
          Gene Collection (MGC), Cancer Genomics Office, National Cancer
          Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892 2590,
          USA
REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
MMENT     Contact: MGC help desk
          Email: cgapbs-r@mail.nih.gov
          Tissue Procurement: ATCC
          cDNA Library Preparation: CLONTECH Laboratories, Inc.
          cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
          DNA Sequencing by: Institute for Systems Biology
          http://www.systemsbiology.org
          contact: amadan@systemsbiology.org
          Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia
          Greene, Mark Ketteman and Anuradha Madan

          Clone distribution: MGC clone distribution information can be found
          through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
          Series: IRAL Plate: 19 Row: 1 Column: 16
          This clone was selected for full length sequencing because it
          passed the following selection criteria: matched mRNA gi: 3327217.
ATURES             Location/Qualifiers
   source          1. .1415
                   /organism="Homo sapiens"
                   /mol_type="mRNA"
```

```
                   /db_xref="taxon:9606"
                   /clone="MGC:42874 IMAGE:4282420"
                   /tissue_type="Testis"
                   /clone_lib="NIH MGC 92"
                   /lab_host="DH10B"
                   /note="Vector: pBluescript"
   CDS             432. .1388
                   /codon_start=1
                   /product="nuclear transcription factor Y, alpha"
                   /protein_id="AAH39244.1"
                   /db_xref="GI:24660184"
                   /db_xref="LocusID:4800"
                   /translation "MEQYTANSNSSTEQIVVQAGQIQQQVQGQPLMVQVSPQLITST
BASE COUNT      417 a    [illegible] c    [illegible] g    [illegible] t
ORIGIN

  Query Match             [illegible]  Score 32;  DB 2;  Length 1415;
  Best Local Similarity   [illegible]  Pred. No. [illegible];
  Matches   32;  Conservative    [illegible];  Mismatches    [illegible];  Indels    0;  Gaps    0;

Qy        [illegible] TGAAAAAAAAAAAAGAAAAAAAAAAAAAAAAA [illegible]
              ||||:|.||||||||||||||||||||||||||
Db        [illegible] TGAAAAAAAAAAAAGAAAAAAAAAAAAAAAAA [illegible]


RESULT 1
BC039244
LOCUS       BC039244                [illegible] bp    mRNA    linear   PRI 06-NOV-2002
DEFINITION  Homo sapiens, [illegible] protein, transcription factor Y, alpha, clone
            MGC:37974 IMAGE:4282420, mRNA, complete cds.
ACCESSION   BC039244
VERSION     BC039244.1  GI:24660184
KEYWORDS    MGC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to [illegible])
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-NOV-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Mikkos Palkovits, M.D., Ph.D.
            cDNA Library Preparation: Michael J. Brownstein (NHGRI, & Soriaki
            Toshiyuki and Piero Carninci (RIKEN)
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Sequencing Group at the Stanford Human Genome
            Center, Stanford University School of Medicine, Stanford, CA 94305
            Web site:       http://www.shgc.stanford.edu
            Contact:  (Dickson, Mark) medspax@lstanford.edu
            Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
            R. M.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 48 Row: i Column: 17
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: [illegible]
FEATURES             Location/Qualifiers
     source          1. .[illegible]
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="MGC:42874 IMAGE:4282420"
                     /tissue_type="Testis"
                     /clone_lib="NIH MGC 92"
                     /lab_host="DH10B"
                     /note="Vector: pBluescript"
     CDS             432. .1388
                     /codon_start=1
                     /product="nuclear transcription factor Y, alpha"
                     /protein_id="AAH39244.1"
                     /db_xref="GI:24660184"
                     /db_xref="LocusID:4800"
                     /translation "MEQYTANSNSSTEQIVVQAGQIQQQVQGQPLMVQVSPQLITST
```

```
                     GQPIMVQAVPGGQGGTIMQVPVSGTQGLQQTQLVPGQIQIQGGQAVQVQGQQGQTQQ
                     IIIQQPQTAVTAGQTQTQQQIAVQGQQVAQTAEGQTIVYQPVNADGTILQQVTVPVSG
                     MITIPAASLAGAQIVQTGANTNTTSSGQGTVTV...VAGNVVNGGGMVMMVPGAGSVP
                     AIQRIPLPGAEMLEEEPLYVNAKQYNRILKRPLSAAKLEAEGKIPKERRKYLHESRHR
                     HAMARKRGEGGRFFSPKEKDSPHMQEPNQAIEEAMIQIIRVS"
BASE COUNT      829 a    745 c    750 g    [illegible] t
ORIGIN

  Query Match           1.0%;  Score 31;  [illegible]
  Best Local Similarity  100.0%;  Pred. No. [illegible]
  Matches   31;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3261 TTGAAAAAAAAAAAAGAAAAAAAAAAAAAAA 3231
             ||||||| | ||||| |||||| |||    |
Db       2972 TTGAAAAAAAAAAAAGAAAAAAAAAAAAAAA 2942


RESULT 1
AT2G18K12
LOCUS       AT2G18K12              [illegible] bp    DNA     linear   PLN 31 JAN 2000
DEFINITION  Arabidopsis thaliana DNA [illegible], BAC clone: T20K12.
ACCESSION   AL137858
VERSION     AL137858.1  GI:6850877
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Mesangiospermae; eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   De Haan,M., Maarse,A.C., Grivell,L.A., Mow,B.W., Lemcke,K.,
            Mayer,K.F.X., Quetier,F. and Salanoubat,M.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 109156)
  AUTHORS   EU Arabidopsis sequencing project.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-JAN 2000) MIPS, at the Max [illegible] for Protein
            Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
            lemcke@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de Project
            coordinator: Marcel Salanoubat and Francis Quetier, Groupement
            d'Interet Public, Centre National de Sequencage - GENOSCOPE, 2 rue
            Gaston Cremieux, BP191, 91006 Evry Cedex, France
            http://www.genoscope.cns.fr
COMMENT     Information on performance of analysis and more detailed
            annotation of this entry and other sequences of chromosome 4, 4
            and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/.
FEATURES             Location/Qualifiers
     source          1..109156
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /variety="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="4"
     gene            125..767
                     /gene="T20K12.10"
     CDS             join(<125..226,521..549,640
                     /gene="T20K12.10"
                     /codon_start=1
                     /product="ribosomal protein S4"
                     /protein_id="CAB71041.1"
                     /db_xref="GI:6850878"
                     /translation="VLNGICLSKQIAELEFRSKVFPFKIGKDNSFFMDVEVQGCFNI
                     TTVFSHSQTVVVCGNKLFILQLFDAKTDLLKREL"
     exon            125..226
                     /gene="T20K12.10"
                     /number=1
     intron          227..520
                     /gene="T20K12.10"
                     /number=1
     exon            521..549
                     /gene="T20K12.10"
                     /number=2
     intron          550..640
                     /gene="T20K12.10"
                     /number=2
     exon            641..767
                     /gene="T20K12.10"
                     /number=3
     repeat_unit     1379..1991
                     /note="inverse repeat of region bp no. 6 to 753"
     repeat_unit     2303..2325
                     /note="23 bp poly(T) tandem repeat"
     gene            6549..8601
                     /gene="T20K12.20"
     gene            complement(join(6549..6636,6750..6898,7025..7060,
                     7151..7192,7295..7394,7490..7551,7723..7798,8420..8601))
                     /gene="T20K12.20"
     CDS             complement(join(6549..6636,6750..6898,7025..7060,
                     7151..7192,7295..7394,7490..7551,7723..7798,8420..8601))
                     /gene="T20K12.20"
                     /note="non-consensus GC donor site at position 7295 and
                     7551, additional T in poly(T) stretch leads to discrepancy
                     of aa sequence to published protein
                     Contains MADS box domain signature and profile
                     AA3-57;Prokaryotic membrane lipoprotein lipid attachment
                     site AA180-190"
                     /codon_start=1
                     /product="MADS-box protein AGL13"
                     /protein_id="CAB71042.1"
                     /db_xref="GI:6850879"
                     /translation="MGRGKVEVKRIENKITRQVTFSKRKSGLLKKAYELSVLCDAEVS
                     LIIFSTGGKLYEFSNVGVGRTIERYYRCKDNLLDNDTLEDTQGLRQEVTKLKCKYESL
                     LRTHRNLVGEDLEGMSIKELQTLERQLEGALSATRKQKTQVMMEQMEELRRKERELGD
                     INNKLKLETEDHDFKGFQDLLLNPVLTAGCSTDFSLQSTHQNYISDCNLGYFLQIGFQ
                     QHYEQGEGSSVTKSNARSDAETNFVQ"
     exon            complement(6549..6636)
                     /gene="T20K12.20"
                     /number=1
     intron          complement(6637..6749)
                     /gene="T20K12.20"
                     /number=1
     exon            complement(6750..6898)
                     /gene="T20K12.20"
                     /number=2
     intron          complement(6899..7024)
                     /gene="T20K12.20"
                     /number=2
     exon            complement(7025..7060)
                     /gene="T20K12.20"
                     /number=3
     intron          complement(7061..7150)
                     /gene="T20K12.20"
                     /number=3
     exon            complement(7151..7192)
                     /gene="T20K12.20"
                     /number=4
     intron          complement(7193..7294)
                     /gene="T20K12.20"
                     /number=4
     exon            complement(7295..7394)
                     /gene="T20K12.20"
                     /number=5
     intron          complement(7395..7489)
                     /gene="T20K12.20"
                     /number=5
     exon            complement(7490..7551)
                     /gene="T20K12.20"
                     /number=6
     intron          complement(7552..7722)
                     /gene="T20K12.20"
                     /number=6
     exon            complement(7723..7798)
                     /gene="T20K12.20"
                     /number=7
     intron          complement(7799..8419)
                     /gene="T20K12.20"
```

```
                /number=7
exon            complement(8420. .8601)
                /gene="T20K12.20"
                /number=8
gene            10534. .13649
                /gene="T20K12.30"
CDS             join(10534. .10684,11012. .11037,11137. .11196,11287. .11412,
                11502. .11565,11643. .11689,11804. .12138,12233. .12590,
                12701. .13279,13374. .13649)
                /gene="T20K12.30"
                /note="similarity to several hypothetical proteins -
                Arabidopsis thaliana"
                /codon_start=1
                /product="putative protein"
                /protein_id="CAB71043.1"
                /db_xref="GI:6850880"
                /translation="MALKRGLSGVNRIRGSGGGSRSVLVLLIFFCVFAPLCFFYSRQV
                YIDSSNDYSIVSVKQNLDWRERLAMQSVRSLFSKEILDVIATSTADAGPLNLDNFKFN
                NLSASWRGTGVDPSFRHSENPATPDVKSNNLNFKRDSISKDSIHQKVETPTKIH RQL
                REKRREMRANELVQHNDDTILKLENAAIERSKSVDSAVLGKYSIWRRENEDNNQSNI
                RLMRDQVIMARVYSGIAKLKNKNDLLQFLQARLKDSQRVLGEATSDADLPRSAIEKLR
                AMGQVLAKAKMQLYDCKLVTGKLRAMLQTADEQVRSLKKQSTFLAQLAAKTIPNFIEC
                LSMRLTIDYYLLSPEKRKFPRSENLENPNLYHYALFSDNVLAASVVVNSTIMNAKDPS
                KHVFHLVTDKLNFGAMNMWFLLNPPGKATIHVENVDEFKWLNSSYCPVLRQLESAAMR
                EYYFKADHPTSGSSNLKYRNPKYLSMLNHLRFYLPEVYPKLNKILFLDDDIVQKDLT
                PLWEVNLNGKVNGAVETCGESFHRFDKYLNFSNPHIARNFNPNACGWAYGMNMFDLKE
                WKKRDITGIYHKWQNMNENRTLWKLGTLPPGLITFYGLTHPLNKAWHVLGLGYNPSID
                KKDIENAAVVHYNGNMKPWLELAMSKYRPYWTKYIKFDHPYLRRCNLHE"
exon            10534. .10684
                /gene="T20K12.30"
                /number=1
intron          10685. .11011
                /gene="T20K12.30"
                /number=1
exon            11012. .11037
                /gene="T20K12.30"
                /number=2
intron          11038. .11136
                /gene="T20K12.30"
                /number=2
exon            11137. .11196
                /gene="T20K12.30"
                /number=3
intron          11197. .11286
                /gene="T20K12.30"
                /number=3
exon            11287. .11412
                /gene="T20K12.30"
                /number=4
intron          11413. .11501
                /gene="T20K12.30"
                /number=4
exon            11502. .11565
                /gene="T20K12.30"
                /number=5
intron          11566. .11642
                /gene="T20K12.30"
                /number=5
exon            11643. .11689
                /gene="T20K12.30"
                /number=6
intron          11690. .11803
                /gene="T20K12.30"
                /number=6
exon            11804. .12138
                /gene="T20K12.30"
                /number=7
intron          12139. .12232
                /gene="T20K12.30"
                /number=7
exon            12233. .12590
                /gene="T20K12.30"
                /number=8
```

```
  Query Match              1  %;  Score 42;  DB 8;  Length 109155;
  Best Local Similarity   10  %;  Pred. No. 2.6e+05;
  Matches   31;  Conservative      3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       6202 TGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 6232
             ||||||||||| | |  |||||||||||||||
Db       6443 TGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 6473


RESULT 12
PFMAL13_00
WPCOMMENT
  Sequence split into 28 fragments  Joins PFMAL13 Accession AL844509
     Fragment Name       Begin       End
     PFMAL13_00          1           [illegible]
     PFMAL13_01          [illegible]
     PFMAL13_02          [illegible]
     PFMAL13_03          [illegible]
     PFMAL13_04          [illegible]
     PFMAL13_05          [illegible]
     PFMAL13_06          [illegible]
     PFMAL13_07          70[illegible]   810[illegible]
     PFMAL13_08          8[illegible]    9[illegible]
     PFMAL13_09          9[illegible]    10[illegible]
     PFMAL13_10          10[illegible]   11[illegible]
     PFMAL13_11          11[illegible]   12[illegible]
     PFMAL13_12          [illegible]     13[illegible]
     PFMAL13_13          13[illegible]   14[illegible]
     PFMAL13_14          14[illegible]   15[illegible]
     PFMAL13_15          15[illegible]   16[illegible]
     PFMAL13_16          16[illegible]   17[illegible]
     PFMAL13_17          17[illegible]   18[illegible]
     PFMAL13_18          18[illegible]   19[illegible]
     PFMAL13_19          19[illegible]   20[illegible]
     PFMAL13_20          20[illegible]   21[illegible]
     PFMAL13_21          21[illegible]   22[illegible]
     PFMAL13_22          22[illegible]   23[illegible]
     PFMAL13_23          23[illegible]   24[illegible]
     PFMAL13_24          24[illegible]   25[illegible]
     PFMAL13_25          25[illegible]   26[illegible]
     PFMAL13_26          26[illegible]   27[illegible]
     PFMAL13_27          27[illegible]   2732359
LOCUS       PFMAL13                 2732359 bp    DNA     linear   HTG 29-JAN-2003
DEFINITION  Plasmodium falciparum 3D7 chromosome 13, *** SEQUENCING IN PROGRESS
            ***, 33 ordered pieces
ACCESSION   AL844509
VERSION     AL844509.1  GI:23495184
KEYWORDS    HTG; HTGS_PHASE2
SOURCE      Plasmodium falciparum 3D7
  ORGANISM  Plasmodium falciparum 3D7
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium
REFERENCE   1  (bases 1 to 2732359)
  AUTHORS   Harris,B., Berriman,M., Clark,L., Lord,A., Barron,A., Corton,C.,
            Berriman,M., Parkhill,J., Hall,N., Atkin,R., Chillingworth,C.,
            Doggett,J., Gwilliam,R., Sanders,M., Hayes,R., Hall,S., Quail,M. and
            Barrell,B.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2732359)
  AUTHORS   Harris,B., Berriman,M., Clark,L., Lord,A., Barron,A., Corton,C.,
            Berriman,M., Parkhill,J., Hall,N., Atkin,R., Chillingworth,C.,
            Doggett,J., Gwilliam,R., Sanders,M., Hayes,R., Hall,S., Quail,M. and
            Barrell,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-SEP-2002) P.falciparum Genome Sequencing Consortium,
            The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA, UK
COMMENT     This chromsome is unfinished and still contains gaps. The contigs
            have been orderd as accurately as possible using mapping data but
            there will be errors. There may also be some sequencing errors.
            Contigs are seperated with 100 Ns.
            For more information about this sequence or the Malaria Project,
            see http://www.sanger.ac.uk/Projects/P_falciparum.
```

```
* NOTE: This is a 'working draft' sequence.  It currently
* consists of 33 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
*         1      77343: contig of 77343 bp in length
*     77344      77443: gap of 100 bp
*     77444      81548: contig of 4105 bp in length
*     81549      81648: gap of 100 bp
*     81649      86880: contig of 5232 bp in length
*     86881      86980: gap of 100 bp
*     86981     372544: contig of 285564 bp in length
*    372545     372644: gap of 100 bp
*    372645     504961: contig of 132317 bp in length
*    504962     505061: gap of 100 bp
*    505062     523311: contig of 18250 bp in length
*    523312     523411: gap of 100 bp
*    523412     539935: contig of 16524 bp in length
*    539936     540035: gap of 100 bp
*    540036     591244: contig of 51209 bp in length
*    591245     591344: gap of 100 bp
*    591345     694818: contig of 103474 bp in length
*    694819     694918: gap of 100 bp
*    694919     709070: contig of 14152 bp in length
*    709071     709170: gap of 100 bp
*    709171     709981: contig of 811 bp in length
*    709982     710081: gap of 100 bp
*    710082     754546: contig of 44465 bp in length
*    754547     754646: gap of 100 bp
*    754647     815774: contig of 61128 bp in length
*    815775     815874: gap of 100 bp
*    815875     832552: contig of 16678 bp in length
*    832553     832652: gap of 100 bp
*    832653     918086: contig of 85434 bp in length
*    918087     918186: gap of 100 bp
*    918187    1128282: contig of 210096 bp in length
*   1128283    1128382: gap of 100 bp
*   1128383    1194073: contig of 65691 bp in length
*   1194074    1194173: gap of 100 bp
*   1194174    1204920: contig of 10747 bp in length
*   1204921    1205020: gap of 100 bp
*   1205021    1236944: contig of 31924 bp in length
*   1236945    1237044: gap of 100 bp
*   1237045    1418883: contig of 181839 bp in length
*   1418884    1418983: gap of 100 bp
*   1418984    1464265: contig of 45282 bp in length
*   1464266    1464365: gap of 100 bp
*   1464366    1580171: contig of 115806 bp in length
*   1580172    1580271: gap of 100 bp
*   1580272    1706827: contig of 126556 bp in length
*   1706828    1706927: gap of 100 bp
*   1706928    1813703: contig of 106776 bp in length
*   1813704    1813803: gap of 100 bp
*   1813804    1822114: contig of 8311 bp in length
*   1822115    1822214: gap of 100 bp
*   1822215    2011820: contig of 189606 bp in length
*   2011821    2011920: gap of 100 bp
*   2011921    2123951: contig of 112031 bp in length
*   2123952    2124051: gap of 100 bp
*   2124052    2238883: contig of 114832 bp in length
*   2238884    2238983: gap of 100 bp
*   2238984    2272314: contig of 33331 bp in length
*   2272315    2272414: gap of 100 bp
*   2272415    2280108: contig of 7694 bp in length
*   2280109    2280208: gap of 100 bp
*   2280209    2571582: contig of 291374 bp in length
*   2571583    2571682: gap of 100 bp
*   2571683    2675082: contig of 103400 bp in length
*   2675083    2675182: gap of 100 bp
*   2675183    2732359: contig of 57177 bp in length.
FEATURES             Location/Qualifiers
     source          1. .2732359
                     /organism="Plasmodium falciparum 3D7"
                     /mol_type="genomic DNA"
                     /isolate="3D7"
                     /db_xref="taxon:36329"
                     /chromosome="13"
     gene            join(22568. .27742,28678. .29991)
                     /gene="VAR"
                     /note="synonyms: MAL13P1.1, PF13_0001"
     CDS             join(22568. .27742,28678. .29991)
                     /gene="VAR"
                     /codon_start=1
                     /product="erythrocyte membrane protein 1 (PfEMP1)"
                     /protein_id="CAD52143.1"
                     /db_xref="GI:23615155"
                     /translation="MGPPGITGTQGETAKHMFDRIGKQVYETVKNEAENYISELEGKL
                     SQATLLGERVSSLKTCQLVEDYRSKANGDVKRYPCANRSPVRFSDESRSQCTYNRIKD
                     NETDDNACGACAPYRRLHLCDYNLEKMGKTSTTKHDLLAEVCMAAKYEGDSIKTHYTI
                     HKHTNNDSAAELCTELARSFADIGDIIRGKDLYLGDIKKKQNGKKTEREKLEENLKRI
                     FGKIHEDVTNGKKEVLKTRYKDINDPEFFKLREDWWTANRATVWKAITCHAGESDKYF
                     RNTCNDSEHSGTFSQPNKYCRCNGDKPGEDKANVDPPTYFDYVPQYLRWFEEWAEDFC
                     RIRKHKLKNAKEQCREKYKSGTDRYCSRNGYDCTQTIRGRNILVSDSECTNCSVVCTP
                     FVKWIENKKLEFEKQKGKYTKEIEKANGTSNGTTIRTQYGTINNMYRKDFYQQLQSGY
                     GDVNAFLELLNKETTCKDHPKVEEKSDIDFNEGTEKTFSHTEYCETCPWCATKKKGID
                     GNWEEQKYEEGCENYLMKPIDESKSTDIDLLVKDTSGTTMVEKLGGLCNDSSKRTVQM
                     ENWQCHYEKKSQYEDGFDKDYCVLKDDKKKKPEHRTIKSYYTLFPNWINEMLKDSIDW
                     SKELKTCINNEKPTNCIRECKSKCDCFKKWVVQKEQEWKQLEEHYEKENFSGDFGPRI
                     SPYVTLEGNLQYSYLEMIRKYYAQEKPVQEIEQIIEKNKNNFEVKEDDNSITKFLQQE
                     KGIATKCIEKQEECKQQKKQQRQKQPADKVVSRSGASPDTASPDTKATEEKEEEEEEE
                     EEEDLGEESEEPEDQAVVDQDGQGETTEKKVPATTEEGSPKETTTPEKSVDVCKTVAE
                     LFSNVDNLKEACTQKYGGNNSRLGWKCIPTSGGEKATGGSGESTGSDATTGGSICVPP
                     RRRRLYVTPLTKWATNMEATEAQAGGDEATEASVPLGDGVSKNPKEALLKAFVESAAV
                     ETFFLWHRYKKIKDKEKKEKEERERENAGLDPFGGSVDDEASTPDPQTQLQSGTIPPD
                     FLRQMFYTLGDYRDICIGGDRDIVGDTIVSNKEGGTPTKISDKIKEILEKVDKKQPDK
                     QNSGTTPKTWWEENGPHIWNAMVCALTYKDNGDKGKPQVDDTVHSQLWDTTKNKPKNR
                     QYEYDQVKLDENSGTGPKSNDPINNPTLKEFISRPPYFRYLEEWGETFCRQRARMLKQ
                     VEKNCTQHGKKQYSGDGEYCDKIHKDPTTLPDLGYSCPKSCSSYRKWIERKKTQYEKQ
                     EKAYNNQKANVQNNNDNGFCGTLEREAAAFLNSLKNGPCKKENGKDNDEYEIKFDDKE
                     KTFKHAKYCDPCPKFKVDCKNGKCDNDKGGDCNGIKTIDAKNFETMVKPTEEIVMRVS
                     DNSGNKFNDLSDCQNAGIFEGIKENKYKCGNVCGYVVCKSENGNGEKVNGIENQNKII
                     TIRGLVAHWVQYFLEDYNKIKHKISHCINNGEVSKCIKDCVKKWVEEKEKEWKKLKEH
                     YQKQYGYNNSGESYPVRSILEQFQSGTEFKNAIKPCGTLQQFESPCGLNGDESSKKEN
                     GNEDAVQCLLKNLGNEAKKCEEKQQTSGIPEAPCVNHTPLEDDDEEPYEDLLLQETEE
                     KPEEAKKNNMPKICKDVVQEAETVVESGCVPAKTPEAPAASPAPKEPPAPSEGTKERP
                     SPPEKKAPLPKEEKTKPKRSLTPTDDPWEPLKNAMLSSTIMWSIGIGFATFTYFYLKK
                     KTKRLVDLFSVINIPKSDYDIPTKLSPNRYIPYTSGKYRGKRYIYLEGDSGTDSGYTD
                     HYSDITSSSFSEYEELDINDIYVPGSPKYKTLIEVVLEPSGNNTTASGKNTPSDTQND
                     IQSDDIPSSKITDNEWNTLKDDFIYNMLQNEPNDIPNDYTSGDIPLNTQPNTLYIDKP
                     LEKPPITSIHDRNLVTGEEYSYNVNMVNSMDDIPINRDNNVYSGIDLINDTLSGNHNV
                     DIYDEVLKRKENELFGINHKKHTTINRFAKPARDDPLHNQLNLPHTWLDRHRDMCEQW
                     NNKEEVLDKLKEEWENETHSGNTHPSDSNKTLNTDVSIQIDMNNPKTTNEFTYVDSNP
                     NQVDDTYVDSNPDNSSMDTILEDLDKPFNEPYYYDMYDDDIYYDVNDDNDISTVDTNA
                     MDVPSKVQIEMDVNTKLVKEKYPISDVWDI"
     gene            complement(join(31809. .32801,33032. .33085))
                     /gene="RIF"
                     /note="synonyms: MAL13P1.2, PF13_0002"
     CDS             complement(join(31809. .32801,33032. .33085))
                     /gene="RIF"
                     /codon_start=1
                     /product="rifin"
                     /protein_id="CAD52144.1"
                     /db_xref="GI:23615156"
                     /translation="MKVHYINILLFAIPLNILIYNQRNHKSTTHHTLKIPITRLLCEC
                     ELYSPVNYDNDPQMKEVMENFNKQTQQRFEEYDERMKTTRQKCKEQCDKEIQKIILKD

  Query Match             1.0%;  Score 31;  DB 2;  Length 110000;
  Best Local Similarity   100.0%;  Pred. No. 2.6e-05;
  Matches   31;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3202 TGAAAAAAAAAAAGAAAAAAAAAAAAAAAAA 3232
             |||||||||||||||||||||||||||||||
```

```
     81649 TGAAAAAAAAAAAGAAAAAAAAAAAAAAAAA 81679

ULT 13
115612
US       AC115612               111882 bp    DNA     linear   INV 12 MAR-2003
INITION  Dictyostelium discoideum chromosome 2 map 6245135 6357017 strain
         AX4, complete sequence.
CESSION  AC115612
RSION    AC115612.2  GI:28828213
YWORDS   HTG.
URCE     Dictyostelium discoideum
RGANISM  Dictyostelium discoideum
         Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
FERENCE  1  (bases 1 to 111882)
AUTHORS  Gloeckner,G., Eichinger,L., Szafranski,K., Pachebat,J., Bean,P.,
         Lehmann,R., Baumgart,C., Parra,G., April,J.F., Guigo,R., Kumpf,K.,
         Tunggal,B., Cox,E., Quail,M.A., Platzer,M., Rosenthal,A. and
         Noegel,A.A.
TITLE    Sequence and analysis of chromosome 2 of Dictyostelium discoideum
JOURNAL  Nature 418 (6893), 79-85 (2002)
MEDLINE  22092622
 PUBMED  12097910
REMARK   The Dictyostelium Genome Sequencing Consortium
FERENCE  2  (bases 1 to 111882)
AUTHORS  Baumgart,C.
TITLE    Direct Submission
JOURNAL  Submitted (21-MAR-2002) Genome Analysis, Institute of Molecular
         Biotechnology, Beutenberstr. 11, Jena 07745, Germany
FERENCE  3  (bases 1 to 111882)
AUTHORS  Baumgart,C.
TITLE    Direct Submission
JOURNAL  Submitted (04-MAR-2003) Genome Analysis, Institute of Molecular
         Biotechnology, Beutenberstr. 11, Jena 07745, Germany
FERENCE  4  (bases 1 to 111882)
AUTHORS  Baumgart,C.
TITLE    Direct Submission
JOURNAL  Submitted (12-MAR-2003) Genome Analysis, Institute of Molecular
         Biotechnology, Beutenberstr. 11, Jena 07745, Germany
MMENT    On Mar 4, 2003 this sequence version replaced gi:19570123.
         CDS predictions from GeneID do not necessarily reflect true genes.
         Further Information is available from IMB Jena, Department of
         Genome Analysis
         (http://genome.imb-jena.de/dictyostelium/)
         and the Univerity Colonge, Institute for Biochemistry I
         (http://www.uni-koeln.de/dictyostelium/project.shtml
         Funding
         Agency : Deutsche Forschungsgemeinschaft (DFG).
ATURES             Location/Qualifiers
     source        1..111882
                   /organism="Dictyostelium discoideum"
                   /mol_type="genomic DNA"
                   /strain="AX4"
                   /db_xref="taxon:44689"
                   /chromosome="2"
                   /map="6245135-6357017"
     CDS           join(23..1814,2160..2164)
                   /note="GeneID exon scores (in order of location ranges):
                   165.11, -1.35 - GSCJ_ID dd_00811"
                   /codon_start=1
                   /product="similar to Plasmodium falciparum. Hypothetical
                   protein"
                   /protein_id="AAL92410.2"
                   /db_xref="GI:28828214"
                   /translation="MHNQQHQQQMHHQHQQQMHQQMHQMHHHHQQHHQPQIQQQQHTS
                   QQQTPQPSPQHTSNAISSVLSMSGSVSSDSIQSPLYEMPHSLLSQQQPQTQPQTQTQT
                   QTQTQAQPQTQPQQHQLQSSNNQINSSSSGIPIPPTHQPQQPSQTQQQNQSTSQQQNQ
                   PTTLQQSQLSPSLNNNVNNPISSPSDDQTPTTQTPSSVVGSPKNKSKKSKTSTTSSST
                   TSTNGTTSTTSNTTLNGASSSSSSTSSSSSNKQKNKTSTNGVGQSPLESSLNSLTKEQ
                   LVQLCTHFLEVEHPDLIQEFENAMPPPDLHDIMQELDDLQKKYNKLYPNNKFCYDNQS
                   YNRVKSGLASFKKALIEYGDNFKEKSGWATLLQYVILAIPYALRMPLWDDEKNNASRI
                   SALSKMDVFGKAAVKGLLQNDIPLEKWERHRNSLHPHSFYLKCTIEEIDKCIEKRRKK
                   EEKSDKQQSKKRKKLNHINGNNTNTTTTTTTNNNNNNSNNNSNSNSNSNSISNNSNN
```

```
                   SNQNNQQQQQQQNNNNNNQNNNNNNNNSSNNMNNKMNNSNNNNSMNNSNNNNIMNQVN
                   MNSQNNIQQQMNNNNMISNQQQNNISMNQSLLQMD"
     CDS           complement(join(2545..4003,4180..5059))
                   /note="GeneID exon scores (in order of location ranges):
                   95.12, -1.09 - GSCJ_ID dd_00813"
                   /codon_start=2
                   /product="similar to Dictyostelium discoideum (Slime
                   mold) 5B protein (Fragment)"
                   /protein_id="AAL92409.2"
                   /db_xref="GI:28828215"
                   /translation="MNITNNNNNNKKDFNDSKCNIHQNEIISFCNNCNNMPLCLKCMC
                   NLESFNSQQMNKGSSELSICYKKIFSIFKDFKNKHLNTLSSFMDLDDFLMQQLEDCM
                   DCLTERFQTIRRINKQVFQLFKCLISIVELDLIKQCNSFLESNRIKCLKNKSELFNDK
                   IVLNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNSDNYCNEENNNDKE
                   NLNFLFTQNQISNFFISPSKYSKFNLFNDFDISFNKESIDLFNSSIKSITTIDIKF
                   FFLFFSQMINFQLQKFLSIQNCNLELIKIPKNIKKLIICDGYKYEIGPGVIPKS
                   LKFLRLDCDSFNPNCLNDSIPSFIEINKTFMMN"
     CDS           complement(<6235..7743)
                   /note="GeneID exon scores (in order of location ranges):
                   18.64 - GSCJ_ID dd_00814"
                   /codon_start=2
                   /product="similar to triacylglyceride lipase cholesterol
                   esterase (Schizosaccharomyces pombe)"
                   /protein_id="AAC50894.1"
                   /db_xref="GI:28828216"
                   /translation="MGLIIVSILSTIWRWSQLYLTLSILIAERITRFFVTTIFGFVPF
                   KYYFDNIISKPVSKDATEAEKDLFRDATEMITARGYPVEEHFVTTPDGFILGLHRIT
                   GPKQFNTIMLSPPTSFRSNNNNNHNHSRSSNNSNGTSSSSTTTTTTTNRKSKLAQKLQ
                   QHVDIDQSTHNSNANTTTTTTTSSGNNNNNNNNMTSTDMMSSASSTSNFSIYNGKPVV
                   VIMHGFMQTSEAWLCRSDPKDSLPFILADAGFDVWLGNNRGNKYSFKHTNYTPNQEGF
                   WNWSLDEIVRYDYPSIVNFICERTGAPSISYIGFSQGTAQGWAALSTNTVISKKINLF
                   IALAPVAFVKGFSNPMIDSLARSRPDFIFLLFGKKSMLPSTIFWRNFLPKDVFVRVID
                   TSVRFLPIWTTENISQQEKKMLYSHIPSYTSVKSVVHWFQIIQTNRFQMFDDILLGIT
                   PNENSKKIIRYHGRVIPQYHPAQILTKCALFYGGSDNLPNTKALLSHLPRDRVVVVHE
                   EPSYEHLIFMWGKDAAEKVFSKVVSLLKSNHFDLTNGTYNGNTNGNNLNNNNCSSNSS
                   PMVNTNNNNYYNSPMLTSQQNNKNNTNNNSIPLKFN"
     CDS           complement(join(9126..9948,10071..10189))
                   /note="GeneID exon scores (in order of location ranges):
                   81.97, -1.57 - GSCJ_ID dd_00816"
                   /codon_start=1
                   /product="similar to methylenetetrahydrofolate
                   dehydrogenase (nad+) (Schizosaccharomyces pombe)"
                   /protein_id="AAC50895.1"
                   /db_xref="GI:28828217"
                   /translation="MEEYKEIMQNPPGKVDVTPIANFFREEVKQQITKNCWKPRVVAF
                   LTVDDQSIDYSKWTKLAVQKIGIEFILKRVERVDLEDLIIEANNDSCVHGILVYYPV
                   FQGMMMTSQDVVSPTKDIEGLSTQNRNLYHNIRFMDGETATKKQVIPCTPDLAMVKI
                   IENLSIVSLAMGRHIKGKTVTIVNRSEIVGRPLAAMLANDGAIVYSIDINGIIPQ
                   QMFSHGIVMSETNVTREEAISKSDILILGVPSPNYKVNSELIQDGTIVINFAGCLNV
                   IPPIQRFLIVPTIGKVTIAMLEPNLLELFNNQISNK"
     CDS           join(<1..98,12369..12477,12564..12652)
                   /note="GeneID exon scores (in order of location ranges):
                   81.64, -1.54 - GSCJ_ID dd_00819"
                   /codon_start=1
                   /product="similar to conserved hypothetical protein
                   [Schizosaccharomyces pombe]"
                   /protein_id="AAL92406.2"
                   /db_xref="GI:28828218"
                   /translation="MIEDKENIERKEGTEVKTTIDINIQSIKDDENGINAAIVTESIT
                   QEISTVSIQQNNSDNNNKLLSSSENSISQLEQLLSKDELIINNEDKTNPLIITTQLDN
                   KNEVEELVVEGEEEDNIKQLTKKEILGVLLVDSFSNFYIQDSILFPQNQTVGVDGEIE
                   KTTTTTGMPTLSNDINLTSPTSIIGNGYKDNEVTSFLVNAFHYLSIPMTVLNYLLLTN
                   KQSISSATISTSTSKQISKVSSFSSITSLLHTPTLTSSTSNSNLLNIVDSSNFEITPT
                   PLSPLISIQQTTITSSTTTTTTAMLKPQKNKENILPNKVDKNIIMELEMLYDNHYNME
                   IFEYSKKISQAQNDIDTKSKLYFQYFSDLTRGSLAICPPKYSDLITSSSQNSSTSNQF
                   YITIRESVEYFLIQNLFHHLFQIDYDLDYQFSKRVESLKFISLSHLGVNKNNSNNGNN
```

```
                SGGSSNQFNKELCPTLLKINVYQTPIDKQRCLAFTLVKLLKINGGEELLLPNLVYLIL
                LSNPPNIYSNYKFLEKFIDHETQNPIYEYSFLLFSMAIQFIEKLDHNHLSIDPLYFQS
                KLFEYNQKNNNNNNNLAATTTNTTTTTTTTTIVNNINNLEISNNNLSVQFSNLLINEP
                DLNNSNNNNEQLNNFLTDNENRFKLFKENNRIQCMIELCKSQITENEIEELLLKFISI
                SLIKENPSYHSLIRYSLKRFFIYYLGIEKNVFLTEMHTIGALQDDNQTNGTQLDLLP
                DGSLANNANNNSSNKDFAIKTAKIAGAAVTGAVIVGMTGGLAAPFIGMMLNFVGAGSI
                VAGITGATGLSSATMLSVIFGAAGAKVSAGAMLATSGIKDYQIHIKKSQTSLHAIIG
                VYGICNDMGGVSDYYQTKSVWERVIRSVTDDYCIFIVEWEKEVMLKLKQLVSEYQQT
                IVQGVIRSAATNIISCSLASALVPLSVIKAASVLNPWSLLLKDRSEKAGKVLAQQIID
                GYFGNRPLTLVSTSMGSRMVFYALRELYNQSLIEKIYSLIESVVFIGSPITSDQKRW
                SKILRLISGRMVNCYTTNDLLKYVCRSINLSTILPAAGVAPIHIPGSSLVIENVD
                LSNLVKSHLDYEKEEILCKILQEIDINNIKYCTFISPLYSGKVYDI"
     CDS        complement(join(15046..16140,16198..16939,17092..17220))
                /note="GeneID exon scores (in order of location ranges):
                140.61, 68.97, 8.44  GSCG ID dd 0922"
                /codon_start=1
                /product="similar to Mus musculus (Mouse). Probable
                urocanate hydratase (EC 4.2.1.49) (Urocamase)
                (Imidazolonepropionate hydrolase)"
                /protein_id="AAC52896.1"
                /db_xref="GI:2884219"
                /translation="MKGLLSSLSKGIPPEPLPEPGGLDSTVPHAPGRPINLTLNEKKL
                AIKNSLRYFPNSLHSILANEFLEELNKYGHLYMYRFRPTHYEMKAYPIEEYPAKSRQA
                ASIMLMIMNNLDREVAQFPNELVTYGGNGSVFQGWAQYHLTMKYLSEMTDQQTLTMYS
                GHPMGLFPSNESSPRVIVTNGMVIPNYSSHKDYGMYAQGVTQYGQMTAGSYCYIGPQ
                GIVHGTTITLLNAGRKYLGVESLIGKVFVSSGLGMSGAQTKAAYICGAICIVAEVSP
                AAVKKRHAQNWVMEKPISIAYQGNVVTLWEKLVEKDITVDLGSDQTSLHNPFNGGYY
                PVQCSFEESNRLMVQEWTKFKEMVQRTLRRHTTAINALSARGMKFWDYGNSFLLEASR
                AKADIIKPGTLDGKEFIYPSYVQDIMGDIESLSPPERWVCTSGNPKDLEITDRIAKEL
                IEKLRSQPGVPKNVQQQFNDNHLWISQAGEHKLVGSQARILYSDCIGRSELAVAPNK
                AVADGTLSAPVVISRDHDVSGTDAPWRETSNIDGSQFCADMAIHNVIGDSFRGATW
                VAIHNGGGTGFGAINAGFGLYLCGSKICEEKAAMMLNWDVNNGIARRSWSYNDNAEN
                TIKRAMELDPSLKVTIPNHTSDFLINSINP"
     CDS        join(18549..18975,19066..19467)
                /note="GeneID exon scores (in order of location ranges):
                55.58, 38.73  GSCG ID dd 00894"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="AAC95067.1"
                /db_xref="GI:2884250"
                /translation="MKISGLLIAASAFASISNLAVLVEEFQYQGGSCSGSQSSIDSQ
                GSFDSQSSWGSQSSWGQSNWGNQSGWASQSGWGSDSSNFLTASESSGEACPSSSIG
                GNIVQEGVCFYYADRLNSIYTVSPDNIVQTFYNLDYNCEGKPICTVTAVSGDCYSAC
                NQGTGITFQLSAVATPVIFTTNTIVEAVYTGECGNYKNSFVSVYYKSLNTCNDNVVA
                CQGTFSFEISCNSTSSAVTIPSNSDCGSPVYGSEPMVNQCGDQYNVIEVCYN"
     CDS        join(19967..20343,20620..20663,21053..21046,21126..21437,
                21609..21804,21908..22038,22393..22502)
                /note="GeneID exon scores (in order of location ranges):
                7.80, 2.74, 5.41, 5.25, 4.55, 2.45, 145.49  GSCG ID
                dd 00805"
                /codon_start=1
                /product="similar to [illegible]
                mold). Hypothetical [illegible]
                /protein_id="AAC[illegible]"
                /db_xref="GI:2884[illegible]"
                /translation="MNRIIILLIIIFFFILVQLIEIKSDNTLYIKNPEKSKSDFC
                SFSFTYILTIFPEIKFETSDSNIIIENDAPLQLIANDCSSLNEYIIPKITFDNYVF
                SVKSNSTSGEKFLSTYFIIPRIVVFLFFIILIELPHTIYQKQVLRIVELNSSCS
                LESIPYYQLQDTGFGYYYIINPDAELNQELVSIVNNSSYTMTIPTIYKNYSNIIGE
```

```
Query Match          1.0%;  Score 31;  DF 2;  [illegible]
Best Local Similarity  100.0%;  Pred. No. [illegible]
Matches   31;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3202 TGAAAAAAAAAAAAGAAAAAAAAAAAAAAAAA 3232
             ||||||||||||||||||||||||||||||||
Db      #2040 TGAAAAAAAAAAAAGAAAAAAAAAAAAAAAAA #2071


RESULT 14
AC012443
LOCUS       AC012443             [illegible]  DNA      linear   HTG 07-NOV-2001
DEFINITION  Homo sapiens BAC clone [illegible]  complete sequence.

ACCESSION   AC012443
VERSION     AC012443.8  GI:16756318
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 153610)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
   PUBMED   9847074
REFERENCE   2  (bases 1 to 153610)
  AUTHORS   Harris,A. and Maupin,R.
  TITLE     The sequence of Homo sapiens BAC clone RP11-17G11
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 153610)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-OCT-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 153610)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-NOV-2001) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Nov 7, 2001 this sequence version replaced gi:13518224.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_NH0017G11
              ------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. John D.
            McPherson, Department of Genetics, Washington University, St. Louis
            MO.  For additional information about the map position of this
            sequence, see http://genome.wustl.edu/gsc

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries. Genomics 51:1-8. The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at the Roswell Park Cancer Institute
            (http://bacpac.med.buffalo.edu)
            VECTOR:  pBACe3.6
            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the left is RP11-288L10; the clone sequenced
            to the right is RP11 150L12.  Actual start of this clone is at base
            position 1 of RP11 17G11; actual end is at base position 153610 of
            RP11-17G11
```

```
         The fidelity of the sequence from base position 96724 to 98766
         cannot be guaranteed due to low quality data and single chemistry
         coverage in an imperfect dinucleotide repeat.
ATURES             Location/Qualifiers
   source          1. .153610
                   /organism="Homo sapiens"
                   /mol_type="genomic DNA"
                   /db_xref="taxon:9606"
                   /chromosome="2"
                   /map="2"
                   /clone="RP11-17G11"
                   /clone_lib="RPCI-11"
   repeat_region   1861. .2149
                   /rpt_family="L1"
   repeat_region   2318. .2645
                   /rpt_family="L1"
   repeat_region   2946. .3401
                   /rpt_family="MaLR"
   repeat_region   3492. .3751
                   /rpt_family="MaLR"
   misc_feature    4131. .4625
                   /note="similar to  EST AA604144 (NID:g2444714) mr69309.s1"
   repeat_region   5019. .5042
                   /rpt_family="(CA)n"
   repeat_region   5043. .5118
                   /rpt_family="MIR"
   repeat_region   6143. .6422
                   /rpt_family="Alu"
   repeat_region   7001. .7263
                   /rpt_family="L1"
   repeat_region   7264. .7570
                   /rpt_family="Alu"
   repeat_region   7571. .8575
                   /rpt_family="L1"
   repeat_region   9999. .10064
                   /rpt_family="L2"
   repeat_region   12325. .12416
                   /rpt_family="L2"
   repeat_region   12550. .12572
                   /rpt_family="(CA)n"
   repeat_region   12574. .12641
                   /rpt_family="MIR"
   repeat_region   12939. .13243
                   /rpt_family="Alu"
   misc_feature    13816. .14113
                   /note="similar to  EST BG687054 (NID:g13918451)"
   misc_feature    13825. .14113
                   /note="similar to  EST BF185866 (NID:g11064370)"
   misc_feature    13826. .14149
                   /note="similar to  EST AV756844 (NID:g10914692)"
   misc_feature    13826. .14130
                   /note="similar to  EST AU136465 (NID:g13997024)"
   misc_feature    13826. .14118
                   /note="similar to  EST BF395435 (NID:g9340800)"
   misc_feature    13826. .14116
                   /note="similar to  EST BF675670 (NID:g11949565)"
   misc_feature    13826. .14113
                   /note="similar to  EST AU133415 (NID:g10933954)"
   misc_feature    13826. .14113
                   /note="similar to  EST BG531665 (NID:g13523202)"
   misc_feature    13826. .14113
                   /note="similar to  EST BG715760 (NID:g13994947)"
   misc_feature    13826. .14095
                   /note="similar to  EST BG107792 (NID:g12601638)"
   misc_feature    13826. .14081
                   /note="similar to  EST BG109801 (NID:g12603307)"
   misc_feature    13839. .14112
                   /note="similar to  EST AU154355 (NID:g11015876)"
   misc_feature    13885. .14113
                   /note="similar to  EST BF215941 (NID:g11109527)"
   misc_feature    13888. .14113
                   /note="similar to  EST BE763113 (NID:g10193037)"
   repeat_region   14546. .14758
```

```
   repeat_region   14[illegible]
   repeat_region   [illegible]
   repeat_region   [illegible]
   repeat_region   [illegible]
   repeat_region   [illegible]
   repeat_region   [illegible]
   repeat_region   [illegible]
   repeat_region   [illegible]
   repeat_region   [illegible]
   misc_feature    [illegible]
   repeat_region   [illegible]
   repeat_region   [illegible]
   repeat_region   [illegible]
   misc_feature    [illegible]
   repeat_region   [illegible]
   repeat_region   [illegible]
   repeat_region   [illegible]
   repeat_region   [illegible]
   repeat_region   [illegible]
   misc_feature    [illegible]
   repeat_region   [illegible]
   repeat_region   [illegible]
   repeat_region   [illegible]
   misc_feature    [illegible]
   repeat_region   [illegible]
   repeat_region   [illegible]
   repeat_region   [illegible]
   repeat_region   [illegible]
   repeat_region   [illegible]

  Query Match                 %;  Score [illegible];  DB 3;  Length 153610;
  Best Local Similarity   [illegible]%;  Pred. No. 2.7e-05;
  Matches   [illegible];  Conservative   [illegible];  Mismatches    0;  Indels    [illegible];  Gaps    0;

Qy       [illegible] TGAAAAAAAAAAAA [illegible] AAAAAAAAAAAAAAAA [illegible]
             ||| ||||||.|| .|| ..||| |||
Db      [illegible] TGAAAAAAAAAAAA [illegible] AAAAAAAAAAAAAAAA [illegible]


RESULT 15
BX469920
LOCUS       BX469920              160584 bp    DNA     linear   HTG 02-JUN-2003
DEFINITION  Danio rerio clone DKEY-237J21, *** SEQUENCING IN PROGRESS ***, 7
            unordered pieces.
ACCESSION   BX469920
VERSION     BX469920.2  GI:[illegible]
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 160594)
  AUTHORS   Mclaren,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-MAY-2003) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: cl erequests@sanger.ac.uk
COMMENT     On May 2, 2003 this sequence version replaced gi:30142544.
              ------- - - Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
               - - ------ Project Information
            Center project name: zK237J21
                 - -------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator, 100% of reads
            Consensus quality: 158650 bases at least Q20
            Consensus quality: 159292 bases at least Q30
            Consensus quality: 159646 bases at least Q40
            Insert size: 159994; sum of contigs
            Insert size: 166857; 4.8% error; agarose fp
            Quality coverage: 5.84x in Q20 bases; sum of contigs Quality
            coverage: 5.74x in Q20 bases; agarose fp
                 - ----.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    26014: contig of 26014 bp in length
            *   26015    26114: gap of 100 bp
            *   26115    55076: contig of 28962 bp in length
            *   55077    55176: gap of 100 bp
            *   55177    93902: contig of 38726 bp in length
            *   93903    94002: gap of 100 bp
            *   94003   123108: contig of 29106 bp in length
            *  123109   123208: gap of 100 bp
            *  123209   140935: contig of 17727 bp in length
            *  140936   141035: gap of 100 bp
            *  141036   158368: contig of 17333 bp in length
            *  158369   158468: gap of 100 bp
            *  158469   160594: contig of 2126 bp in length.
FEATURES             Location/Qualifiers
     source          1..160594
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-237J21"
                     /clone_lib="DanioKey"
     misc_feature    1..26014
                     /note="assembly fragment:01012
                     fragment_chain:1"
     misc_feature    26115..55076
                     /note="assembly fragment:00855
                     fragment_chain:1"
     misc_feature    55177..93902
                     /note="assembly fragment:02954
                     fragment_chain:1"
     misc_feature    94003..123108
                     /note="assembly fragment:00217
                     fragment_chain:1"
     misc_feature    123209..140935
                     /note="assembly fragment:00779
                     fragment_chain:2"
     misc_feature    141036..158368
                     /note="assembly fragment:00250
                     fragment_chain:2"
     misc_feature    158469..160594
                     /note="assembly_fragment:01497.0"
BASE COUNT    50133 a  28321 c  28491 g  53048 t    601 others
ORIGIN

  Query Match             1.0%;  Score 31;  DB 2;  Length 160594;
  Best Local Similarity   100.0%;  Pred. No. 2.7e-05;
  Matches   31;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3202 TGAAAAAAAAAAAGAAAAAAAAAAAAAAAAA 3232
              ||||||||.|||||||||  ||||||||||||
Db     150127 TGAAAAAAAAAAAGAAAAAAAAAAAAAAAAA 150157


Search completed: October 31, 2003, 05:47:02
Job time : 11516.5 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 30, 2003, 16:29:58 ; Search time 785.73 Seconds
(without alignments)
11104.788 Million cell updates/sec

Title: US-09-980-043-1
Perfect score: 3232
Sequence: 1 gtcgacccacgcgtccgt[illegible] 3232

Scoring table: OLIGO NUC
Gapop 60.0 , Gapext 60.0

Searched: 2552756 seqs, 13497199017 residues

Word size : 0

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqn-em:1/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-em:1/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-em:1/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-em:1/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-em:1/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-em:1/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-em:1/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-em:1/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-em:1/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-em:1/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-em:1/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-em:1/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-em:1/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-em:1/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-em:1/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-em:1/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-em:1/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-em:1/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-em:1/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-em:1/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-em:1/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-em:1/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-em:1/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-em:1/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseqn-em:1/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	3232	100.0	3232	22	AAC65448	Populus tremuloide
2	172	5.3	1010	22	AAC65449	Populus tremuloide
3	35	1.1	3207	19	AAV98372	Cellulose synthase
4	32	1.0	3747	21	AAA87145	Eucalyptus grandis
c 5	30	0.9	127	20	AAX24845	Infectious HCV gen
c 6	30	0.9	176	20	AAX24849	Infectious HCV gen
c 7	30	0.9	183	20	AAX24846	Infectious HCV gen
c 8	30	0.9	201	20	AAX24850	Infectious HCV gen
c 9	30	0.9	225	20	AAX24844	Infectious hepatit
c 10	30	0.9	245	22	AAF23493	3' UTR of Hepatiti
c 11	30	0.9	309	24	ABN93954	Gene #452 used to
12	30	0.9	348	22	AAI82665	Human polynucleoti
c 13	30	0.9	473	22	AAL25797	Human breast cance
c 14	30	0.9	558	22	AAL18963	Human breast cance
15	30	0.9	1131	20	AAX84964	Human secreted pro
16	30	0.9	2667	21	AAA78382	Human secreted pro
c 17	30	0.9	4167	22	ABA08224	Human ovarian and
c 18	30	0.9	4167	22	AAS29236	Genomic sequence #
19	30	0.9	4167	22	AAS30018	Human lung antigen
c 20	30	0.9	4167	22	AAL05022	Human reproductive
c 21	30	0.9	4167	22	AAL07543	Human reproductive
22	30	0.9	4167	22	AAS28711	Genomic sequence #
c 23	30	0.9	4167	22	AAK68040	Human immune/haema
24	30	0.9	4167	22	AAK89382	Human digestive sy
c 25	30	0.9	4167	23	ABL97915	Human testicular a
c 26	30	0.9	4167	24	ABS68376	Human DNA-binding
c 27	30	0.9	9518	22	AAD03778	Hepatitis C virus
c 28	30	0.9	9518	22	AAD03807	Hepatitis C virus
c 29	30	0.9	9518	22	AAD03808	Hepatitis C virus
c 30	30	0.9	9595	20	AAX24843	Infectious hepatit
c 31	30	0.9	9595	22	AAC86939	Nucleotide sequenc
c 32	30	0.9	9595	22	AAF23492	Infectious Hepatit
c 33	30	0.9	9599	20	AAX24832	Infectious hepatit
c 34	30	0.9	9599	20	AAX24833	Infectious hepatit
c 35	30	0.9	9599	22	AAC86938	Nucleotide sequenc
c 36	30	0.9	9599	22	AAF23491	Infectious Hepatit
c 37	30	0.9	9611	22	AAC86645	Nucleotide sequenc
c 38	30	0.9	9611	22	AAC86646	Nucleotide sequenc
c 39	30	0.9	9611	22	AAC86647	Nucleotide sequenc
c 40	30	0.9	9611	22	AAC86648	Nucleotide sequenc
c 41	30	0.9	10803	25	ABX10617	MKO-Z nucleotide s
42	29	0.9	64	25	ABX50113	Bovine EST associa
43	29	0.9	67	25	ABX50316	Bovine EST associa
44	29	0.9	76	25	ABX41011	Bovine EST associa
c 45	29	0.9	78	24	ABV96830	Human pancreatic c

ALIGNMENTS

RESULT 1
AAC65448
ID AAC65448 standard; cDNA; 3232 BP.
XX
AC AAC65448;
XX
DT 07-MAR-2001 (first entry)
XX
DE Populus tremuloides CelA cDNA.
XX
KW Aspen; CelA; cellulose synthase; arabidopsis; wood;
KW mechanical stress response element; MSRE; cellulose biosynthesis; ss.
XX
OS Populus tremuloides.
XX
PN WO200071670-A2.
XX
PD 30-NOV-2000.
XX
PF 18-MAY-2000; 2000WO-US13637.
XX
PR 21-MAY-1999; 99US-0135280.
XX
PA (UNMI) UNIV MICHIGAN TECHNOLOGICAL.
XX
PI Chiang VL, Wu L, Chandrashekhar JP, Carraway DT;
XX
DR WPI; 2001-032025/04.
DR P-PSDB; AAB37893.
XX
PT Novel polynucleotide encoding cellulose synthase useful for enhancing

cellulose biosynthesis, modifying lignin biosynthesis, and for accelerating growth of plants and for upregulating cellulose synthase levels -

Claim 1; Fig 1; 65pp; English.

The present sequence is given in a specification relating to polynucleotides encoding cellulose synthase enzymes from Populus tremuloides and Arabidopsis. The cellulose synthases are designated PtCelA and AraxCelA respectively. The polynucleotides may be introduced into plants where they are useful for altering the growth of a plant. Vectors comprising a cellulose synthase promoter, or a fragment of the promoter containing one or more mechanical stress response elements (MSRE), operably linked to a cellulose synthase gene are useful for causing stress-induced gene expression in a plant cell. A cDNA encoding a protein that binds to a positive MSRE may be delivered into the plant to increase expression of cellulose synthase. A cDNA in an antisense orientation may be delivered to reduce expression of cellulose synthase. The polynucleotides enable the development of transgenic tree varieties having increased cellulose content, decreased lignin content and therefore improved wood fibre characteristics. The cellulose synthase promoters are useful for synthesising greater quantities of high-crystalline cellulose in plant and preferably in trees, which permits an increase in proportion of cellulose in transgenic plants, greater strength of juvenile wood and fiber and acceleration of overall growth rate.

Sequence 3232 BP; 895 A; 668 C; 772 G; 897 T; 0 other;

Query Match 100.0%; Score 3232; DB 22; Length 3232;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3232; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    1 GTCGACCCACGCGTCCGTCTTGAAAGAATATGAAGTTGTAAAGAGCTGGTAAAGTGGTAA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 GTCGACCCACGCGTCCGTCTTGAAAGAATATGAAGTTGTAAAGAGCTGGTAAAGTGGTAA 60

Qy   61 TAAGCAAGATGATGGAATCTGGGGCTCCTATATGCCATACCTGTGGTGAACAGGTGGGGC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 TAAGCAAGATGATGGAATCTGGGGCTCCTATATGCCATACCTGTGGTGAACAGGTGGGGC 120

Qy  121 ATGATGCAAATGGGGAGCTATTTGTGGCTTGCCATGAGTGTAGCTATCCCATGTGCAAGT 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 ATGATGCAAATGGGGAGCTATTTGTGGCTTGCCATGAGTGTAGCTATCCCATGTGCAAGT 180

Qy  181 CTTGTTTCGAGTTTGAAATCAATGAGGGCCGGAAAGTTTGCTTGCGGTGTGGCTCGCCAT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 CTTGTTTCGAGTTTGAAATCAATGAGGGCCGGAAAGTTTGCTTGCGGTGTGGCTCGCCAT 240

Qy  241 ATGATGAGAACTTGCTGGATGATGTAGAAAAGAAGGGGTCTGGCAATCAATCCACAATGG 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 ATGATGAGAACTTGCTGGATGATGTAGAAAAGAAGGGGTCTGGCAATCAATCCACAATGG 300

Qy  301 CATCTCACCTCAACGATTCTCAGGATGTCGGAATCCATGCTAGACATATCAGTAGTGTGT 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 CATCTCACCTCAACGATTCTCAGGATGTCGGAATCCATGCTAGACATATCAGTAGTGTGT 360

Qy  361 CCACTGTGGATAGTGAAATGAATGATGAATATGGGAATCCAATTTGGAAGAATCGGGTGA 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 CCACTGTGGATAGTGAAATGAATGATGAATATGGGAATCCAATTTGGAAGAATCGGGTGA 420

Qy  421 AGAGCTGTAAGGATAAAGAGAACAAGAAGAAAAAAGAGAAGTCCTAAGGCTGAAACTGAAC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 AGAGCTGTAAGGATAAAGAGAACAAGAAGAAAAAAGAGAAGTCCTAAGGCTGAAACTGAAC 480

Qy  481 CAGCTCAAGTTCCTACAGAACAGCAGATGGAAGAGAAACCGTCTGCAGAGGCTTCGGAGC 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481 CAGCTCAAGTTCCTACAGAACAGCAGATGGAAGAGAAACCGTCTGCAGAGGCTTCGGAGC 540

Qy  541 CGCTTTCAATTGTTTATCCAATTCCACGCAACAAGCTCACACCATACAGAGCAGTGATCA 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541 CGCTTTCAATTGTTTATCCAATTCCACGCAACAAGCTCACACCATACAGAGCAGTGATCA 600

Qy  601 TTATGCGACTGTCTTTTGGGTCTTTCTTCACTTCAGAATAACAAATCCTGTCGATA 660
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601 TTATGCGACTGTCTTTTGGGTCTTTCTTCACTTCAGAATAACAAATCCTGTCGATA 660

Qy  661 GTGCGTTTCGCCTTTATTTCTTCATAGTGAGATCTGGTTTGCATTTTCTTGGG 720
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  661 GTGCGTTTCGCCTTTATTTCTTCATAGTGAGATCTGGTTTGCATTTTCTTGGG 720

Qy  721 TGTTGGATGAGTTAAAGTTTAAATTGTCAATAGAGAAACGTATATCGAAAGGCTGT 780
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  721 TGTTGGATGAGTTAAAGTTTAAATTGTCAATAGAGAAACGTATATCGAAAGGCTGT 780

Qy  781 CGGCAAGCGTATTAAAAAGAGGTTAATTTCTTAGCTTGCTGGTGTGGATTTTTTCGTGA 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  781 CGGCAAGCGTATTAAAAAGAGGTTAATTTCTTAGCTTGCTGGTGTGGATTTTTTCGTGA 840

Qy  841 GTACTGTTAATTTTGAAGAAGAAGTTAATTGATCACTGCCAATACAGTGCTTTTCATCC 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  841 GTACTGTTAATTTTGAAGAAGAAGTTAATTGATCACTGCCAATACAGTGCTTTTCATCC 900

Qy  901 TTGTTTTTGAATTTGATGGTTATAGGTTCGAGTTA 960
          |||||||||||||||||||||||||||||||||||
Db  901 TTGTTTTTGAATTTGATGGTTATAGGTTCGAGTTA 960

Qy  961 TGCTTTTCAATTTAATTGTTGTAAAAGAATTTGCAAGGAAGTGGGTTTCGTTCT 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  961 TGCTTTTCAATTTAATTGTTGTAAAAGAATTTGCAAGGAAGTGGGTTTCGTTCT 1020

Qy 1021 GCAAAAAATTCTTTGAAGTAAAATTTTACTTCTCACAGAAAATTGATT 1080
          |||||||||||||||||||||||||||||||||||||||||||||||
Db 1021 GCAAAAAATTCTTTGAAGTAAAATTTTACTTCTCACAGAAAATTGATT 1080

Qy 1081 ACTTGAAAGACAAATTCAACCTTTTTTGTTAAAGAACGTAGAGCAATGAAAAGGGATT 1140
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1081 ACTTGAAAGACAAATTCAACCTTTTTTGTTAAAGAACGTAGAGCAATGAAAAGGGATT 1140

Qy 1141 ATGAAGAGTACAAATCCGAGTTAATCCCTTTAGCAAAGGCTCAGAAAACACCTGAAG 1200
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1141 ATGAAGAGTACAAATCCGAGTTAATCCCTTTAGCAAAGGCTCAGAAAACACCTGAAG 1200

Qy 1201 AAGGATGGACTATCAAGATGGAACACTTGGCGTGGGAATAACACACGTGATCACCTTG 1260
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1201 AAGGATGGACTATCAAGATGGAACACTTGGCGTGGGAATAACACACGTGATCACCTTG 1260

Qy 1261 GGCATGATTCAATTTCTTGTGGAAATACTCTGAGCTCGTGACATTGAAGGAAATGAAC 1320
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1261 GGCATGATTCAATTTCTTGTGGAAATACTCTGAGCTCGTGACATTGAAGGAAATGAAC 1320

Qy 1321 TACCTCGTCTACTATGTCTCCACGGAGAAGAGACCTGGCTACCAGCACCACAAAAAGG 1380
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1321 TACCTCGTCTACTATGTCTCCACGGAGAAGAGACCTGGCTACCAGCACCACAAAAAGG 1380

Qy 1381 CTGGTGCAGAAAATTCTCTGGTGAGAGTGTCTTCAGTACTCACAAATGCTCCCTACATCC 1440
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1381 CTGGTGCAGAAAATTCTCTGGTGAGAGTGTCTTCAGTACTCACAAATGCTCCCTACATCC 1440

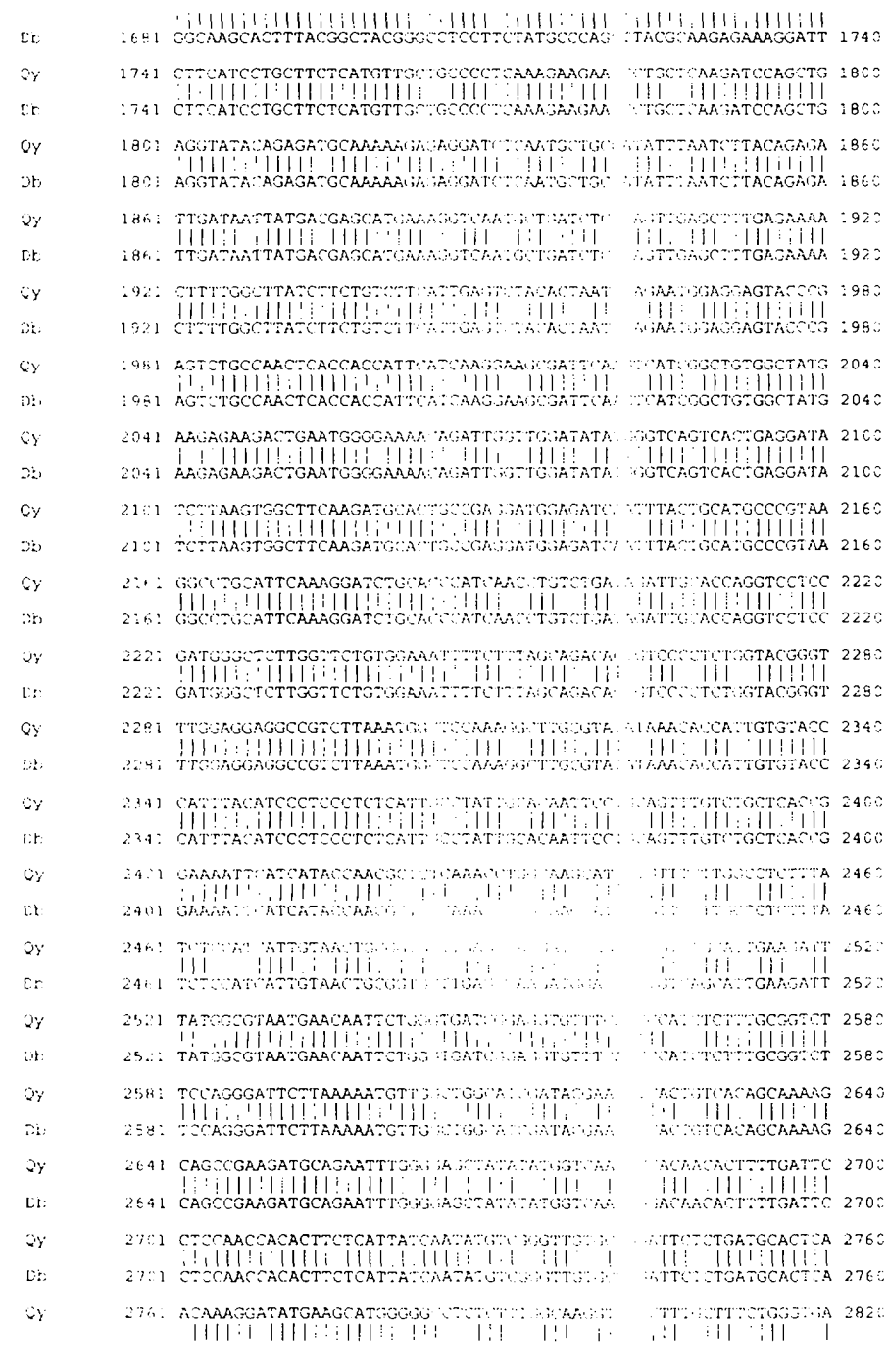
Qy 1441 TCAATGTTGATTTCATCACTATGTAAACAATACTAAGGCTGTTCGAGAGGCAATGTGCA 1500
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1441 TCAATGTTGATTTCATCACTATGTAAACAATACTAAGGCTGTTCGAGAGGCAATGTGCA 1500

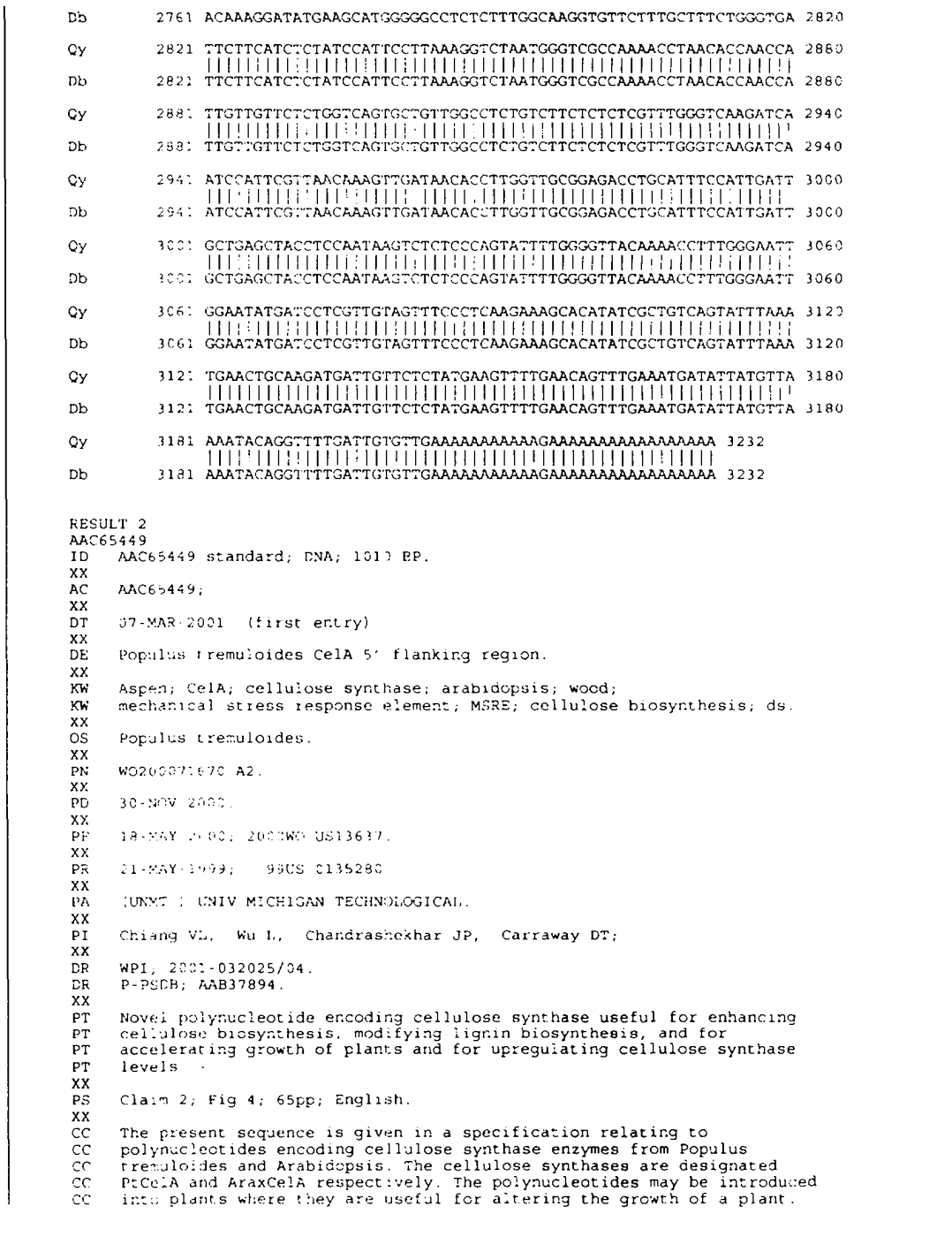
Qy 1501 TCCTGATGGACCCAAAGTAGGTCGAGATGTATGCTATGTGCAGTTCCCTCAGAGGTTTG 1560
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1501 TCCTGATGGACCCAAAGTAGGTCGAGATGTATGCTATGTGCAGTTCCCTCAGAGGTTTG 1560

Qy 1561 ATGGCATAGATAAGAGTGATCGCTACGCCAATCGTAACGTAGTTTTCTTTGATGTTAACA 1620
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1561 ATGGCATAGATAAGAGTGATCGCTACGCCAATCGTAACGTAGTTTTCTTTGATGTTAACA 1620

Qy 1621 TGAAAGGGTTGGATTGGCATTCAAGGACCAGTATACGTAGGAACTGGTTGTGTTTTCAACA 1680
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1621 TGAAAGGGTTGGATTGGCATTCAAGGACCAGTATACGTAGGAACTGGTTGTGTTTTCAACA 1680

Qy 1681 GGCAAGCACTTTACGGCTACGGGCCTCCTTCTATGCCCAGCTTACGCAAGAGAAAGGATT 1740
```





RESULT 2
AAC65449
ID AAC65449 standard; DNA; 1013 BP.
XX
AC AAC65449;
XX
DT 07-MAR-2001 (first entry)
XX
DE Populus tremuloides CelA 5' flanking region.
XX
KW Aspen; CelA; cellulose synthase; arabidopsis; wood;
KW mechanical stress response element; MSRE; cellulose biosynthesis; ds.
XX
OS Populus tremuloides.
XX
PN WO200071670 A2.
XX
PD 30-NOV 2000.
XX
PF 18-MAY 2000; 2000WO US13637.
XX
PR 21-MAY-1999; 99US 0135280
XX
PA (UNMT) UNIV MICHIGAN TECHNOLOGICAL.
XX
PI Chiang VL, Wu L, Chandrashekhar JP, Carraway DT;
XX
DR WPI; 2001-032025/04.
DR P-PSDB; AAB37894.
XX
PT Novel polynucleotide encoding cellulose synthase useful for enhancing
PT cellulose biosynthesis, modifying lignin biosynthesis, and for
PT accelerating growth of plants and for upregulating cellulose synthase
PT levels -
XX
PS Claim 2; Fig 4; 65pp; English.
XX
CC The present sequence is given in a specification relating to
CC polynucleotides encoding cellulose synthase enzymes from Populus
CC tremuloides and Arabidopsis. The cellulose synthases are designated
CC PtCelA and AraxCelA respectively. The polynucleotides may be introduced
CC into plants where they are useful for altering the growth of a plant.

Vectors comprising a cellulose synthase promoter, or a fragment of the promoter containing one or more mechanical stress response elements (MSRE), operably linked to a cellulose synthase gene are useful for causing stress-induced gene expression in a plant cell. A cDNA encoding a protein that binds to a positive MSRE may be delivered into the plant to increase expression of cellulose synthase. A cDNA in an antisense orientation may be delivered to reduce expression of cellulose synthase. The polynucleotides enable the development of transgenic tree varieties having increased cellulose content, decreased lignin content and therefore improved wood fibre characteristics. The cellulose synthase promoters are useful for synthesising greater quantities of high crystalline cellulose in plant and preferably in trees, which permits an increase in proportion of cellulose in transgenic plants, greater strength of juvenile wood and fiber and acceleration of overall growth rate.

Sequence 1010 BP; 289 A; 212 C; 216 G; 293 T; 0 other;

Query Match 5.3%; Score 172; DB 22; Length 1010;
Best Local Similarity 100.0%; Pred. No. 1.1e-57;
Matches 172; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      31 TGAAGTTGTAAAGAGCTGGTAAAGTGGTAATAAGCAAGATGATGGAATCTGGGGCTCCTA 90
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     803 TGAAGTTGTAAAGAGCTGGTAAAGTGGTAATAAGCAAGATGATGGAATCTGGGGCTCCTA 862

Qy      91 TATGCCATACCTGTGGTGAACAGGTGGGGCATGATGCAAATGGGGAGCTATTTGTGGCTT 150
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     863 TATGCCATACCTGTGGTGAACAGGTGGGGCATGATGCAAATGGGGAGCTATTTGTGGCTT 922

Qy     151 GCCATGAGTGTAGCTATCCCATGTGCAAGTCTTGTTTCGAGTTTGAAATCAA 202
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     923 GCCATGAGTGTAGCTATCCCATGTGCAAGTCTTGTTTCGAGTTTGAAATCAA 974
```

RESULT 3
AAV08372
AAV08372 standard; cDNA to mRNA; 3207 BP.

AAV08372;

08-FEB-1999 (first entry)

Cellulose synthase PcsA1 coding sequence.

Cellulose synthase; cotton; ds.

Gossypium hirsutum.

Key	Location/Qualifiers
CDS	77..3001
	/*tag= a

EP875575-A2.

04-NOV 1998

31 MAR-1998; 98EP-0302489.

01-APR-1997; 97JP-0083133.

(NISN) NISSHINBO IND INC.

Aotsuka S, Hasegawa O, Hayashi T, Ihara Y;

WPI; 1998-559440/48.
P-PSDB; AAW73308.

New DNA encoding cotton cellulose synthase proteins - useful for controlling cellulose synthesis in prokaryotic or eukaryotic cells

Example 5; Page 14-18; 51pp; English.

CC This sequence encodes a cellulose synthase of the invention. The DNA can
CC be used for contr[illegible] cellulose synthesis in a cell by introducing the
CC DNA into the cell, exp[illegible] RNA with a nucleotide sequence homologous
CC to the DNA or a nucleic [illegible] sequence complementary to the DNA. The cells
CC are prokaryotic cells [illegible] Acetobacterium and/or eukaryotic cells such as
CC yeasts i.e. Saccharo[illegible] plant cells (cotton plants) and cultured
CC cells of mammals etc
XX
SQ Sequence 3207 BP; 884 A; [illegible] C; [illegible] G; 849 T; 0 other;

Query Match [illegible]; Score 60; DB 19; Length 3207;
Best Local Similarity [illegible] %; Pred. No. 1.00054;
Matches 35; Conservative [illegible]; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1136 GCAAAGTTTAAGAAGTCTTGTTTCGAGTTTGAAATCAA 1215
             ||||||||||||||||||||||||||||||||||||||||
Db      1147 GCAAAGTT[illegible]                     1184
```

RESULT 4
AAA67145
ID AAA67145 standard; DNA; [illegible] BP.
XX
AC AAA67145;
XX
DT 31-OCT 2000 (first entry)
XX
DE Eucalyptus grandis cellulose synthase nucleotide sequence SEQ ID NO:141.
XX
KW Eucalyptus grandis; [illegible] Monterey pine; modification;
KW plant cell wall; polysaccharide; polysaccharide biosynthetic pathway;
KW transgenic plant; wo
XX
OS Eucalyptus grandis
XX
PN WO200062092-A2.
XX
PD 20-APR-2000
XX
PF 08-OCT 1999; 99WO-NZ 00[illegible]
XX
PR 13-OCT 1998; 98US-0103462.
PR 11-AUG 1999; 99US-0374426.
XX
PA (GENE) GENESIS RES & DEV CORP LTD.
PA (FLET) FLETCHER CHALLEN OF FORESTS LTD.
XX
PI Bloksberg LN;
XX
DR WPI; 2000-339328/29.
DR P-PSDB; AAB16338.
XX
PT New genes encoding proteins involved in a plant polysaccharide
PT biosynthetic pathway, useful for modulation or altering the
PT polysaccharide content, composition or structure of the plant
XX
PS Claim 1; Page 196-107; 410pp; English.
XX
CC The present invention describes isolated polynucleotides (PN) comprising
CC a sequence selected [illegible] one of 835 nucleotide sequences given in
CC AAA67073 to AAA67907, their (reverse) complements, sequences producing
CC an Expectation (E) value of 0.01 or less compared to the 835 sequences,
CC sequences at least 50% identical to them, 200, 100, 40 or 20-mers of the
CC 835 sequences or sequences that are degenerately equivalent or allelic
CC to the 835 sequences. The polynucleotides are used to modify the
CC activity of a polypeptide involved in a polysaccharide biosynthetic
CC pathway in the plant. They are especially used to modulate or alter the
CC polysaccharide content, composition or structure of the plant. AAB16268
CC to AAB16340 are proteins encoded by some of the polynucleotide sequence
CC given in the present invention.
XX
SQ Sequence 3747 BP; 976 A; 822 C; 935 G; 1014 T; 0 other;

Query Match 1.0%; Score 32; DB 21; Length 3747;
Best Local Similarity 100.0%; Pred. No. 0.008;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       2811 TTCTGGGTGATTCTTCATCTCTATCCATTCCT 2842
             ||||||||||||||||||||||||||||||||
Db       3256 TTCTGGGTGATTCTTCATCTCTATCCATTCCT 3287
```

RESULT 5
AAX24845/c
ID AAX24845 standard; DNA; 127 BP.
XX
AC AAX24845;
XX
DT 21-JUN-1999 (first entry)
XX
DE Infectious HCV genotype 1a clone pCV-H77C(-98X) 3'UTR.
XX
KW HCV; infectious clone; infection; diagnosis; therapy; vaccine;
KW screening; assay; antiviral; virucide; ss.
XX
OS Hepatitis C virus.
XX
FH Key Location/Qualifiers
FT CDS 1..3
FT /*tag= a
FT /partial
FT /note= "stop codon of coding region"
FT 3'UTR 4..127
FT /*tag= b
FT /partial
XX
PN WO9904008-A2.
XX
PD 28-JAN-1999
XX
PF 16-JUL-1998; 98WO-US14688
XX
PR 27-JAN-1998; 98US-0014416.
PR 18-JUL-1997; 97US-0053062.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Bukh J, Emerson SU, Purcell RH, Yanagi M;
XX
DR WPI; 1999-132252/11.
XX
PT New isolated hepatitis C virus nucleic acids - used to develop
PT products for the diagnosis, prevention and treatment of HCV
PT infections and for developing screening assays
XX
PS Example 10; Fig 17A; 126pp; English.
XX
CC The present sequence comprises the 3' untranslated region (3'UTR)
CC of a hepatitis C virus (HCV) genotype 1a infectious clone pCV-H77C
CC deletion mutant, termed pCV-H77C(-98X), in which the 98 most 3'
CC nucleotides of pCV-H77C (see AAX24844) have been deleted. Seven
CC deletion mutants (see AAX24845-51) of pCV-H77C have been created
CC that have various deletions in the 3'UTR. Construct pCV-H77C(-98X)
CC was tested for infectivity in chimpanzees. The results indicated
CC that the 3' 98 nucleotides of the 3'UTR are critical for production
CC of infectious HCV. The invention discloses nucleic acid sequences
CC (see AAX24832-33 and AAX24843) which encode infectious HCV, and the use
CC of these sequences, and polypeptides (see AAW98020-22) encoded by
CC them, in the development of vaccines and diagnostics for HCV and in
CC the development of screening assays for the identification of
CC antiviral agents for HCV.
XX
SQ Sequence 127 BP; 11 A; 25 C; 11 G; 80 T; 0 other;

Query Match 0.9%; Score 30; DB 20; Length 127;
Best Local Similarity 100.0%; Pred. No. 0.077;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       3203 GAAAAAAAAAAAGAAAAAAAAAAAAAAAAA 3232
              ||||||||||||||||||||||||||||||
Db         84 GAAAAAAAAAAAGAAAAAAAAAAAAAAAAA 55
```

RESULT 6
AAX24849/c
ID AAX24849 standard; DNA; 176 BP.
XX
AC AAX24849;
XX
DT 21-JUN-1999 (first entry)
XX
DE Infectious HCV genotype 1a clone pCV-H77C(+49X) 3'UTR.
XX
KW HCV; infectious clone; infection; diagnosis; therapy; vaccine;
KW screening; assay; antiviral; virucide; ss.
XX
OS Hepatitis C virus.
XX
FH Key Location/Qualifiers
FT CDS 1..3
FT /*tag= a
FT /partial
FT /note= "stop codon of coding region"
FT 3'UTR 4..176
FT /*tag= b
FT /partial
XX
PN WO9904008-A2.
XX
PD 28-JAN-1999.
XX
PF 16-JUL-1998; 98WO-US14688.
XX
PR 27-JAN-1998; 98US-0014416.
PR 18-JUL-1997; 97US-0053062.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Bukh J, Emerson SU, Purcell RH, Yanagi M;
XX
DR WPI, 1999-132252/11.
XX
PT New isolated hepatitis C virus nucleic acids - used to develop
PT products for the diagnosis, prevention and treatment of HCV
PT infections and for developing screening assays
XX
PS Example 10; Fig 17B; 126pp, English.
XX
CC The present sequence comprises the 3' untranslated region (3'UTR)
CC of a hepatitis C virus (HCV) genotype 1a infectious clone pCV-H77C
CC deletion mutant, termed pCV-H77C(+49X), in which the proximal 49
CC nucleotides of the 3' conserved region of the 3'UTR of pCV-H77C
CC (see AAX24844) have been removed. 7 Deletion mutants (see AAX24845-51)
CC of pCV-H77C have been created that have various deletions in the 3'UTR.
CC Construct pCV-H77C(-98X) (see AAX24845) was tested for infectivity in
CC chimpanzees. The results indicated that the 3' 98 nucleotides of
CC the 3'UTR are critical for production of infectious HCV. The
CC invention discloses nucleic acid sequences (see AAX24832-33 and
CC AAX24843) which encode infectious HCV, and the use of these sequences,
CC and polypeptides (see AAW98020-22) encoded by them, in the development
CC of vaccines and diagnostics for HCV and in the development of
CC screening assays for the identification of antiviral agents for HCV.
XX
SQ Sequence 176 BP; 21 A; 38 C; 25 G; 92 T; 0 other;

Query Match 0.9%; Score 30; DB 20; Length 176;
Best Local Similarity 100.0%; Pred. No. 0.068;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
      3203 GAAAAAAAAAAAGAAAAAAAAAAAAAAAAA 3232
           ||||||||||||||||||||||||||||||||
        84 GAAAAAAAAAAAGAAAAAAAAAAAAAAAAA 55
```

SULT 7
X24846/c
AAX24846 standard; DNA; 183 BP.

AAX24846;

21-JUN-1999 (first entry)

Infectious HCV genotype 1a clone pCV-H77C(-42X) 3'UTR.

HCV; infectious clone; infection; diagnosis; therapy; vaccine; screening; assay; antiviral; virucide; ss.

Hepatitis C virus.

Key	Location/Qualifiers
CDS	1..3
	/*tag= a
	/partial
	/note= "stop codon of coding region"
3'UTR	4..183
	/*tag= b
	/partial

WO9904008-A2.

28-JAN-1999.

16-JUL-1998; 98WO-US14688.

27-JAN-1998; 98US-0014416.
18-JUL-1997; 97US-0053062.

(USSH) US DEPT HEALTH & HUMAN SERVICES.

Bukh J, Emerson SU, Purcell RH, Yanagi M;

WPI; 1999-132252/11.

New isolated hepatitis C virus nucleic acids - used to develop products for the diagnosis, prevention and treatment of HCV infections and for developing screening assays

Example 10; Fig 17A; 126pp; English.

The present sequence comprises the 3' untranslated region (3'UTR) of a hepatitis C virus (HCV) genotype 1a infectious clone pCV-H77C deletion mutant, termed pCV-H77C(-42X), in which the 42 most 3' nucleotides of pCV-H77C (see AAX24844) have been deleted. Seven deletion mutants (see AAX24845-51) of pCV-H77C have been created that have various deletions in the 3'UTR. Construct pCV-H77C(-98X) (see AAX24845) was tested for infectivity in chimpanzees. The results indicated that the 3' 98 nucleotides of the 3'UTR are critical for production of infectious HCV. The invention discloses nucleic acid sequences (see AAX24832-33 and AAX24843) which encode infectious HCV, and the use of these sequences, and polypeptides (see AAW98020-22) encoded by them, in the development of vaccines and diagnostics for HCV and in the development of screening assays for the identification of antiviral agents for HCV.

Sequence 183 BP; 21 A; 41 C; 28 G; 93 T; 0 other;

```
Query Match             0.9%;  Score 30;  DB 20;  Length 183;
Best Local Similarity   100.0%;  Pred. No. 0.067;
Matches   30;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

      3203 GAAAAAAAAAAAGAAAAAAAAAAAAAAAAA 3232
           ||||||||||||||||||||||||||||||||
```

```
Db         84 GAAAAAAAAAAAGAAAAAAAAAAAAAAAAA 55
```

RESULT 8
AAX24850/c
ID AAX24850 standard; DNA [illegible] BP.
XX
AC AAX24850;
XX
DT 21-JUN-1999 (first [illegible])
XX
DE Infectious HCV genotype [illegible] clone pCV-H77C(-VE-24) 3'UTR.
XX
KW HCV; infectious clone; [illegible]; diagnosis; therapy; vaccine;
KW screening; assay; [illegible].
XX
OS Hepatitis C virus.
XX
FH Key [illegible]
FT CDS [illegible]
FT /*tag=
FT /partial
FT /note= "stop codon of coding region"
FT 3'UTR 4..201
FT /*tag= b
FT /partial
XX
PN WO9904008-A2
XX
PD 28-JAN-1999.
XX
PF 16-JUL-1998; 98WO-US14688.
XX
PR 27-JAN-1998; 98US-0014416.
PR 18-JUL-1997; 97US-0053062.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Bukh J, Emerson SU, Purcell RH, Yanagi M.
XX
DR WPI; 1999-132252/11.
XX
PT New isolated hepatitis C virus nucleic acids - used to develop
PT products for the diagnosis, prevention and treatment of HCV
PT infections and for developing screening assays
XX
PS Example 10; Fig 17B; 126pp; English.
XX
CC The present sequence comprises the 3' untranslated region (3'UTR)
CC of a hepatitis C virus (HCV) genotype 1a infectious clone pCV-H77C
CC deletion mutant, termed pCV-H77C(-VE-24), in which the first 24
CC nucleotides of the 3' variable region of the 3'UTR of pCV-H77C
CC (see AAX24844) have been removed. 7 Deletion mutants (see AAX24845-51)
CC of pCV-H77C have been created that have various deletions in the 3'UTR.
CC Construct pCV-H77C(-98X) (see AAX24845) was tested for infectivity in
CC chimpanzees. The results indicated that the 3' 98 nucleotides of
CC the 3'UTR are critical for production of infectious HCV. The
CC invention discloses nucleic acid sequences (see AAX24832-33 and
CC AAX24843) which encode infectious HCV, and the use of these sequences,
CC and polypeptides (see AAW98020-22) encoded by them, in the development
CC of vaccines and diagnostics for HCV and in the development of
CC screening assays for the identification of antiviral agents for HCV.
XX
SQ Sequence 201 BP; 25 A; [illegible] C; 32 G; 93 T; 0 other;

```
  Query Match             0.9%;  Score 30;  DB 20;  Length 201;
  Best Local Similarity   100.0%;  Pred. No. 0.067;
  Matches   30;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3203 GAAAAAAAAAAAGAAAAAAAAAAAAAAAAA 3232
             ||||||||||:|||.||||||||||||||||
Db         60 GAAAAAAAAAAAGAAAAAAAAAAAAAAAAA 31
```

```
RESULT 9
AAX24844/c
ID   AAX24844 standard; DNA; 225 BP.
XX
AC   AAX24844;
XX
DT   21-JUN-1999  (first entry)
XX
DE   Infectious hepatitis C virus genotype 1a clone pCV-H77C 3'UTR.
XX
KW   HCV; infectious clone; infection; diagnosis; therapy; vaccine;
KW   screening; assay; antiviral; virucide; ss.
XX
OS   Hepatitis C virus.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..3
FT                   /*tag= a
FT                   /partial
FT                   /note= "stop codon of coding region"
FT   3'UTR           4..225
FT                   /*tag= b
XX
PN   WO9904008-A2.
XX
PD   28-JAN-1999
XX
PF   16-JUL-1998;   98WO-US14688.
XX
PR   27-JAN-1998;   98US-0014416.
PR   18-JUL-1997;   97US-0053062.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Bukh J;  Emerson SU;  Purcell RH;  Yanagi M;
XX
DR   WPI; 1999-132252/11.
XX
PT   New isolated hepatitis C virus nucleic acids - useful to develop
PT   products for the diagnosis, prevention and treatment of HCV
PT   infections and for developing screening assays
XX
PS   Example 10; Fig 17A; 126pp; English.
XX
CC   The present sequence comprises the 3' untranslated region (3'UTR)
CC   of hepatitis C virus (HCV) genotype 1a infectious clone pCV-H77C,
CC   including the coding region stop codon, a 44 base 3'-terminal variable
CC   region, an 81 nucleotide poly(U/UC) tract and a 98 nucleotide 3'
CC   conserved region. Seven deletion mutants (see AAX24845-51) of
CC   pCV-H77C have been created that have various deletions in the
CC   3'UTR. Construct pCV-H77C(-98X) (see AAX24849) was tested for
CC   infectivity in chimpanzees. The results indicate that the 3' 98
CC   nucleotides of the 3'UTR (conserved region) were critical for
CC   production of infectious HCV. The invention discloses nucleic acid
CC   sequences (see AAX24832-33 and AAX24847) which encode infectious HCV,
CC   and the use of these sequences, and polypeptides (see AAW98020-22)
CC   encoded by them, in the development of vaccines and diagnostics for
CC   HCV and in the development of screening assays for the identification
CC   of antiviral agents for HCV.
XX
SQ   Sequence 225 BP; 30 A; 51 C; 40 G; 104 T; 0 other;

  Query Match             0.9%;  Score 30;  DB 20;  Length 225;
  Best Local Similarity   100.0%;  Pred. No. 0.066;
  Matches   30;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3203 GAAAAAAAAAAAAGAAAAAAAAAAAAAAAAA 3232
             ||||||||||||||||||||||||||||||
Db         84 GAAAAAAAAAAAAGAAAAAAAAAAAAAAAAA 55


RESULT 10
AAF23493/c
ID   AAF23493 standard; DNA; 245 BP.
XX
AC   AAF23493;
XX
DT   21-MAR-2001  (first entry)
XX
DE   3' UTR of Hepatitis C virus.
XX
KW   GBV B; hepatitis C virus; HCV; vaccine; ss.
XX
OS   Hepatitis C virus.
XX
PN   WO200075337-A1.
XX
PD   14-DEC-2000.
XX
PF   02-JUN-2000; 2000WO-US15293.
XX
PR   04-JUN-1999;   99US-0137694.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Bukh J,  Yanagi M,  Emerson SU,  Purcell RH;
XX
DR   WPI; 2001-091214/10.
XX
PT   New infectious nucleic acids of the GB virus-B clone, useful for
PT   indirectly studying the molecular properties of hepatitis C virus (HCV)
PT   and in developing vaccines and therapeutics for HCV -
XX
PS   Disclosure; Fig 4; 96pp; English.
XX
CC   The present invention relates to GB virus-B. The nucleic acid molecules
CC   of the invention are useful for indirectly studying the molecular
CC   properties of hepatitis C virus (HCV). The infectious nucleic acid
CC   sequence of the GB virus-B clone and the HCV/GBV-B chimeras may be used
CC   in the development of vaccines and therapeutics for HCV.
XX
SQ   Sequence 245 BP; 36 A; 62 C; 41 G; 106 T; 0 other;

  Query Match             0.9%;  Score 30;  DB 22;  Length 245;
  Best Local Similarity   100.0%;  Pred. No. 0.065;
  Matches   30;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3203 GAAAAAAAAAAAAGAAAAAAAAAAAAAAAAA 3232
             ||||||||||||||||||||||||||||||
Db        104 GAAAAAAAAAAAAGAAAAAAAAAAAAAAAAA 75


RESULT 11
ABN93954/c
ID   ABN93954 standard; DNA; 309 BP.
XX
AC   ABN93954;
XX
DT   13-AUG-2002  (first entry)
XX
DE   Gene #452 used to diagnose liver cancer.
XX
KW   Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;
KW   metastatic liver tumour; cytostatic; expression profile; disease state;
KW   disease progression; drug toxicity; drug efficacy; drug metabolism.
XX
OS   Homo sapiens.
XX
PN   WO200229103-A2.
XX
PD   11-APR-2002.
XX
PF   02-OCT-2001; 2001WO-US30589.
XX
PR   02-OCT-2000; 2000US-237054P.
```

(GENE-) GENE LOGIC INC.

Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;

WPI; 2002-426119/45.

Diagnosing and detecting the progression of liver cancer, hepatocellular carcinoma or metastatic liver tumor in a patient, involves detecting the level of expression of two or more genes in a liver tissue sample -

Claim 1; SEQ ID NO 452; 298pp; English.

The invention relates to a novel method for diagnosing and detecting the progression of liver cancer, hepatocellular carcinoma or metastatic liver tumour in a patient, and differentiating metastatic liver cancer from hepatocellular carcinoma in a patient, involving detecting the level of expression of two or more genes represented in ABN93503 ABN97455 in a tissue sample. The method of the invention has hepatotropic, and cytostatic activity. The method is useful for diagnosing and detecting the progression of liver cancer, hepatocellular carcinoma and metastatic liver carcinoma in a patient. The method is useful for identifying expression profiles which serve as useful diagnostic markers as well as markers that can be used to monitor disease states, disease progression, drug toxicity, drug efficacy and drug metabolism.
Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

Sequence 309 BP; 100 A; 45 C; 45 G; 119 T; 0 other;

Query Match 0.9%; Score 30; DB 24; Length 309;
Best Local Similarity 100.0%; Pred. No. 0.064;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       3202 TGAAAAAAAAAAAGAAAAAAAAAAAAAAAAA 3231
             ||||| ||||||:|||||||:|||||||||
Db         30 TGAAAAAAAAAAAGAAAAAAAAAAAAAAAAA 1
```

SULT 12
82665
AAI82665 standard; cDNA; 348 BP.

AAI82665;

06-NOV-2001 (first entry)

Human polynucleotide SEQ ID NO 2725.

Human; cytokine; cell proliferation; cell differentiation; gene therapy; vaccine; peptide therapy; stem cell growth factor; haematopoiesis; tissue growth factor; immunomodulatory; cancer; leukaemia; nervous system disorders; arthritis; inflammation; ss.

Homo sapiens.

WO200164835-A2.

07-SEP-2001.

26-FEB-2001; 2001WO-US04927.

28-FEB-2000; 2000US-0515126.
18-MAY-2000; 2000US-0577409.

(HYSE-) HYSEQ INC.

Tang YT, Liu C, Drmanac RT;

WPI; 2001-514838/56.
P-PSDB; AAO02734.

XX
PT Isolated nucleic acids [illegible] polypeptides, useful for preventing
PT diagnosing and treating [illegible] inflammation and immune
PT disorders
XX
PS Claim 1; SEQ ID NO 2725; [illegible]pp; English.
XX
CC The invention relates to [illegible] polynucleotides (AAI79941 AAI03641) and
CC the encoded proteins (AAI [illegible] AAI03616) that exhibit activity relating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in stem cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine like activities,
CC e.g. stem cell [illegible] haematopoiesis regulating
CC activity, tissue growth [illegible] activity, immunomodulatory activity and
CC activin/inhibin [illegible] in the diagnosis and/or
CC treatment of [illegible] disorders, arthritis and
CC inflammation
CC Note: The sequence data [illegible] for part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 348 BP; 174 A; [illegible]; 58 G; [illegible] other;

Query Match [illegible]; Score 30; DB 22; Length 348;
Best Local Similarity [illegible]; Pred. No. [illegible];
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       3203 GAAAAAAAAAAAGAAAAAAAAAAAAAAAAA 3232
             ||||| |||||:|||||||:|||||||||
Db        209 GAAAAAAAAAAAGAAAAAAAAAAAAAAAAA 238
```

RESULT 13
AAI25797/c
ID AAI25797 standard; cDNA; 474 BP.
XX
AC AAI25797;
XX
DT 07-DEC-2001 (first entry)
XX
DE Human breast cancer expressed polynucleotide 18254
XX
KW Human; breast cancer; [illegible] marker, cytostatic; ss.
XX
OS Homo sapiens.
XX
PN WO200151628-A2.
XX
PD 19-JUL-2001.
XX
PF 10-JAN-2001; 2001WO-US00798.
XX
PR 14-JAN-2000; 2000US-0176477
PR 14-MAR-2000; 2000US-0189469
PR 24-MAR-2000; 2000US-0191793
PR 29-MAR-2000; 2000US-0192468.
PR 15-MAY-2000; 2000US-0571210.
PR 09-JUN-2000; 2000US-0211415.
PR 25-JUL-2000; 2000US-0622[illegible]84.
XX
PA (MILL-) MILLENNIUM PHARMACEUTICALS INC.
XX
PI Lillie J, Xu Y, Wang Y, Steinmann K;
XX
DR WPI; 2001-451856/48.
XX
PT New peptide useful as a marker for the diagnosis of breast cancer
XX
PS Claim 1; Page 3374-3375; 3695pp; English.
XX
CC The invention relates to human breast cancer expressed polynucleotides
CC (AAL07544-AAL26789) and methods of assessing whether a patient is

```
CC   afflicted with breast cancer by examining the correlation between the
CC   expression of certain markers and the cancerous state of breast cells.
CC   The polynucleotides and encoded polypeptides are potential markers for
CC   detecting, diagnosing, monitoring, characterising treating and
CC   potentially preventing breast cancer. The polynucleotides and encoded
CC   polypeptides are also useful for isolating compounds with cytostatic
CC   activity.
XX
SQ   Sequence 473 BP; 125 A; 90 C; 101 G; 157 T; 0 other;

  Query Match              0.9%;  Score 30;  DB 22;  Length 473;
  Best Local Similarity   100.0%;  Pred. No. 0.41;
  Matches   30;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3203 GAAAAAAAAAAAGAAAAAAAAAAAAAAAAA 3232
             ||||||||||||||||||||||||||||||
Db        286 GAAAAAAAAAAAGAAAAAAAAAAAAAAAAA 315


RESULT 14
AAL18963/c
ID   AAL18963 standard; cDNA; 558 BP.
XX
AC   AAL18963;
XX
DT   07-DEC-2001  (first entry)
XX
DE   Human breast cancer expressed polynucleotide #14.
XX
KW   Human; breast cancer; cell marker; cytostatic; ds
XX
OS   Homo sapiens
XX
PN   WO200151620-A2
XX
PD   19-JUL-2001
XX
PF   10-JAN-2001; 2001WO-US00798.
XX
PR   14-JAN-2000; 2000US-0176077.
PR   14-MAR-2000; 2000US-0189167.
PR   24-MAR-2000; 2000US-0192099.
PR   29-MAR-2000; 2000US-0193480.
PR   15-MAY-2000; 2000US-0205230.
PR   09-JUN-2000; 2000US-0211315.
PR   25-JUL-2000; 2000US-0220934.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Lillie J, Xu Y, Wang Y, [illegible];
XX
DR   WPI; 2001-451470-48.
XX
PT   New peptide useful as a marker [illegible] breast cancer
XX
PS   Claim 1; Page 2041-2042; 3695pp; English.
XX
CC   The invention relates to human breast cancer [illegible] polynucleotides
CC   (AAL07544-AAL26789) and methods of assessing whether a patient is
CC   afflicted with breast cancer by examining the correlation between the
CC   expression of certain markers and the cancerous state of breast cells.
CC   The polynucleotides and encoded polypeptides are potential markers for
CC   detecting, diagnosing, monitoring, characterising treating and
CC   potentially preventing breast cancer. The polynucleotides and encoded
CC   polypeptides are also useful for isolating compounds with cytostatic
CC   activity.
XX
SQ   Sequence 558 BP; 152 A; 109 C; 116 G; 181 T; [illegible];

  Query Match              0.9%;  Score 30;  DB 22;  Length 558;
  Best Local Similarity   100.0%;  Pred. No. [illegible];
  Matches   30;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3203 GAAAAAAAAAAAGAAAAAAAAAAAAAAAAA 3232
             ||||||||||||||||||||||||||||||
Db        296 GAAAAAAAAAAAGAAAAAAAAAAAAAAAAA 267


RESULT 15
AAX84964
ID   AAX84964 standard; DNA; 1131 BP.
XX
AC   AAX84964;
XX
DT   30-JUL-1999  (first entry)
XX
DE   Human secreted protein gene No. 32.
XX
KW   Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW   diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW   developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
KW   immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW   inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KW   cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW   osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW   endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
XX
OS   Homo sapiens.
XX
PN   WO9924836-A1.
XX
PD   20-MAY-1999.
XX
PF   04-NOV-1998;   98WO-US23435.
XX
PR   17-NOV-1997;   97US-0066100.
PR   07-NOV-1997;   97US-0064900.
PR   07-NOV-1997;   97US-0064908.
PR   07-NOV-1997;   97US-0064911.
PR   07-NOV-1997;   97US-0064912.
PR   07-NOV-1997;   97US-0064983.
PR   07-NOV-1997;   97US-0064984.
PR   07-NOV-1997;   97US-0064985.
PR   07-NOV-1997;   97US-0064987.
PR   07-NOV-1997;   97US-0064988.
PR   17-NOV-1997;   97US-0066090.
PR   17-NOV-1997;   97US-0066094.
PR   17-NOV-1997;   97US-0066095.
PR   17-NOV-1997;   97US-0066089.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Carter KC,  Ebner R,  Endress GA,  Feng P,  Janat F;
PI   Kyaw H,  Lafleur DW,  Moore PA,  Ni J,  Olsen HS,  Rosen CA;
PI   Rubinfeld SM,  Shi Y,  Soppet DR,  Wei Y;
XX
DR   WPI; 1999-337740/28.
DR   P-PSDB; AAY27598.
XX
PT   New human secreted proteins and coding sequences useful for treating
PT   disorders of the immune system and hyperproliferative disorders
XX
PS   Claim 1; Page 298; 507pp; English.
XX
CC   This sequence represents a nucleic acid molecule which encodes a
CC   secreted human protein. The gene number is given in the descriptor line.
CC   The gene can be used to generate fusion proteins by linking to the gene
CC   to a human immunoglobulin Fc portion (e.g. AAX84924) for increasing the
CC   stability of the fused protein as compared to the human protein only.
CC   The invention relates to 125 novel genes and their fragments (nucleic
CC   acid sequences: AAX84933-X85057; amino acid sequences AAY27567-Y27933)
CC   which are useful for preventing, treating or ameliorating medical
CC   conditions e.g. by protein or gene therapy. Also, pathological
CC   conditions can be diagnosed by determining the amount of the new
CC   polypeptides in a sample or by determining the presence of mutations in
CC   the new polynucleotides. Specific uses are described for each of the 125
```

polynucleotides, based on which tissues they are most highly expressed in (see AAX84933 for described uses).

Sequence 1131 BP; 341 A; 214 C; 251 G; 323 T; 2 other;

uery Match 0.9%; Score 30; DB 20; Length 1131;
est Local Similarity 100.0%; Pred. No. 0.055;
atches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
      3203 GAAAAAAAAAAAGAAAAAAAAAAAAAAAAA 3232
           ||||||||||||||||||||||||||||||
      1070 GAAAAAAAAAAAGAAAAAAAAAAAAAAAAA 1099
```

rch completed: October 30, 2003, 23:10:32
time : 788.73 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 30, 2003, 21:03:32 ; Search time 6343.45 Seconds
(without alignments)
12387.170 Million cell updates/sec

Title: US-09-980-043-1
Perfect score: 3232
Sequence: 1 gtcgacccacgcgtccgtcc ... aaaaaaaaaaaaaaa 3232

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 22781392 seqs, 12150248858 residues

Word size: 0

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	418	12.9	484	12	BI128887	BI128887 G083P04Y
2	267	8.3	457	12	BI131891	BI131891 G127P18Y
3	245	7.6	479	12	BI132826	BI132826 G135P19Y
4	229	7.1	571	13	BU895617	BU895617 X028D07 P
c 5	146	4.5	496	9	AI166604	AI166604 xylem.est
6	135	4.2	476	13	BU897587	BU897587 X060B01 P
7	135	4.2	648	13	BU894292	BU894292 X005H01 P
8	135	4.2	683	13	BU897715	BU897715 X067B10 P
9	131	4.1	289	12	BI128439	BI128439 G076P02Y
10	123	3.8	287	12	BI129538	BI129538 G092P15Y
11	114	3.5	434	13	BU894130	BU894130 X003C01 P
12	98	3.0	435	12	BI127228	BI127228 G057P27Y
13	93	2.9	126	12	BI129892	BI129892 G097P01Y
14	82	2.5	337	13	BU897069	BU897069 X050E11 P
15	77	2.4	233	13	BU896785	BU896785 X045F11 P
16	46	1.4	466	13	BU898166	BU898166 X075H06 P
17	46	1.4	475	13	BU895665	BU895665 X029B04 P
18	46	1.4	479	12	BI127210	BI127210 G004P82Y
c 19	46	1.4	483	12	BI130322	BI130322 G103P84Y
c 20	44	1.4	606	14	CB819948	CB819948 EST 940 G
21	36	1.1	463	12	BI131187	BI131187 G116P69Y
22	34	1.1	218	13	BQ388280	BQ388280 NISC_mn28
c 23	34	1.1	482	13	BQ396766	BQ396766 NISC_ng22
24	32	1.0	158	13	BU531948	BU531948 AGENCOURT
25	32	1.0	197	9	AW486670	AW486670 75666 MAR
26	32	1.0	285	9	AI163338	AI163338 A040P09U
27	32	1.0	438	9	AL673551	AL673551 AL673551
28	32	1.0	577	14	CA753427	CA753427 002200030
c 29	32	1.0	1464	10	BF312119	BF312119 601898007
30	31	1.0	58	2	HSM003270	Al038794 Homo sapi
c 31	31	1.0	64	2	HSM003271	Al038795 Homo sapi
32	31	1.0	102	14	CD522259	CD522259 AGENCOURT
c 33	31	1.0	194	14	CD417338	CD417338 Gm_ck7952
34	31	1.0	241	13	BU556115	BU556115 AGENCOURT
35	31	1.0	241	14	CA779650	CA779650 MPL384_1_
36	31	1.0	265	13	BU960653	BU960653 AGENCOURT
37	31	1.0	320	9	AI251181	AI251181 qv37e08.x
38	31	1.0	361	13	BY675335	BY675335 BY675335
39	31	1.0	458	10	BG168868	BG168868 602320194
40	31	1.0	467	10	BF130532	BF130532 601818674
c 41	31	1.0	468	9	AW333291	AW333291 S19G12 AG
42	31	1.0	475	12	BI380731	BI380731 BFLG1_002
c 43	31	1.0	480	14	CB979193	CB979193 CAB40007_
44	31	1.0	485	13	BU704657	BU704657 UI-M-FP0
c 45	31	1.0	500	9	AA568210	AA568210 nf13h11.s

ALIGNMENTS

RESULT 1
BI128887
LOCUS BI128887 484 bp mRNA linear EST 31-DEC-2001
DEFINITION G083P04Y Populus cambium cDNA library Populus tremula x Populus tremuloides cDNA, mRNA sequence.
ACCESSION BI128887
VERSION BI128887.1 GI:18012858
KEYWORDS EST.
SOURCE Populus tremula x Populus tremuloides
ORGANISM Populus tremula x Populus tremuloides
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids ; eurosids I; Malpighiales; Salicaceae; Populus.
REFERENCE 1 (bases 1 to 484)
AUTHORS Hertzberg,M., Aspeborg,H., Erlandsson,R., Bjorkbacka,H., Hiltonen ,T., Karlsson,J., Teeri,T., Gustafsson,P., Bahlerao,R., Jansson,S., Nilsson,O., Sundberg,B., Nilsson,P., Uhlen,M., Sandberg,G. and Lundeberg,J.
TITLE Gene expression in Populus
JOURNAL Unpublished
COMMENT Contact: Erlandsson R
Department of Biotechnology
Royal Institute of Technology
Teknikringen 30, Stockholm S-10044, Sweden
Tel: 46 8 790 8287
Fax: 46 8 245452
Email: rikerl@biochem.kth.se.

```
TURES             Location/Qualifiers
  source          1. .484
                  /organism="Populus tremula x Populus tremuloides"
                  /mol_type="mRNA"
                  /db_xref="taxon:47664"
                  /clone_lib="Populus cambium cDNA library"
                  /note="Organ: cambium"
E COUNT      122 a    115 c    101 g    146 t
GIN

uery Match            12.9%;  Score 418;  DB 12;  Length 484;
est Local Similarity  100.0%;  Pred. No. 3e-74;
atches  418;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

      2290 GCCGTCTTAAATGGCTCCAAAGGCTTGCGTATATAAACACCATTGTGTACCCATTTACAT 2349
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
         1 GCCGTCTTAAATGGCTCCAAAGGCTTGCGTATATAAACACCATTGTGTACCCATTTACAT   60

      2350 CCCTCCCTCTCATTGCCTATTGCACAATTCCTGCAGTTTGTCTGCTCACCGGAAAATTCA 2409
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
        61 CCCTCCCTCTCATTGCCTATTGCACAATTCCTGCAGTTTGTCTGCTCACCGGAAAATTCA 120

      2410 TCATACCAACGCTCTCAAACCTGGCAAGCATGCTGTTTCTTGGCCTCTTTATCTCCATCA 2469
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
       121 TCATACCAACGCTCTCAAACCTGGCAAGCATGCTGTTTCTTGGCCTCTTTATCTCCATCA 180

      2470 TTGTAACTGCGGTGCTTGAGCTAAGATGGAGCGGTGTCAGCATTGAAGATTTATGGCGTA 2529
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
       181 TTGTAACTGCGGTGCTTGAGCTAAGATGGAGCGGTGTCAGCATTGAAGATTTATGGCGTA 240

      2530 ATGAACAATTCTGGGTGATCGGAGGTGTTTCAGCCCATCTCTTTGCGGTCTTCCAGGGAT 2589
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
       241 ATGAACAATTCTGGGTGATCGGAGGTGTTTCAGCCCATCTCTTTGCGGTCTTCCAGGGAT 300

      2590 TCTTAAAAATGTTGGCTGGCATCGATACGAACTTCACTGTCACAGCAAAAGCAGCCGAAG 2649
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
       301 TCTTAAAAATGTTGGCTGGCATCGATACGAACTTCACTGTCACAGCAAAAGCAGCCGAAG 360

      2650 ATGCAGAATTTGGGGAGCTATATATGGTCAAGTGGACAACACTTTTGATTCCTCCAAC 2707
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
       361 ATGCAGAATTTGGGGAGCTATATATGGTCAAGTGGACAACACTTTTGATTCCTCCAAC 418


SULT 2
 31891
'US        BI131891                 457 bp    mRNA    linear   EST 31-DEC-2001
TINITION   G127P18Y Populus cambium cDNA library Populus tremula x Populus
           tremuloides cDNA, mRNA sequence.
CESSION    BI131891
RSION      BI131891.1  GI:18015862
YWORDS     EST.
URCE       Populus tremula x Populus tremuloides
ORGANISM   Populus tremula x Populus tremuloides
           Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
           Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
           ; eurosids I; Malpighiales; Salicaceae; Populus.
FERENCE    1  (bases 1 to 457)
AUTHORS    Hertzberg,M., Aspeborg,H., Erlandsson,R., Bjorkbacka,H., Hiltonen
           ,T., Karlsson,J., Teeri,T., Gustafsson,P., Bahlerao,R., Jansson,S.,
           Nilsson,O., Sundberg,B., Nilsson,P., Uhlen,M., Sandberg,G. and
           Lundeberg,J.
TITLE      Gene expression in Populus
JOURNAL    Unpublished
MMENT      Contact: Erlandsson R
           Department of Biotechnology
           Royal Institute of Technology
           Teknikringen 30, Stockholm S-10044, Sweden
           Tel: 46 8 790 8287
           Fax: 46 8 245452
           Email: rikerl@biochem.kth.se.
ATURES             Location/Qualifiers
    source         1. .457
                   /organism="Populus tremula x Populus tremuloides"
                   /mol_type="mRNA"
                   /db_xref="taxon:47664"
                   /clone_lib="Populus cambium cDNA library"
                   /note="Organ: cambium"
BASE COUNT      105 a    114 c    134 g    104 t
ORIGIN

  Query Match             [illegible]  Score 267;  DB 12;  Length 457;
  Best Local Similarity   [illegible]  Pred. No. 7.4e-44;
  Matches  267;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2264 TCCTCTCTCATTGCCTATTGCACAATTCCTGCAGTTTGTCTGCTCACCGGAAAATTCA 2323
             ||| ||| ||| ||| ||| ||| ||| ||| |||||||||||||||||||||||||||
Db        138 TCCTCTCTCATTGCCTATTGCACAATTCCTGCAGTTTGTCTGCTCACCGGAAAATTCA 197

Qy       2324 AAACACCAA[illegible]                                        2383
             || ||| |                                                   |||
Db        198 AAACACCAA[illegible]                                        257

Qy       2384 AGTTTGTCTTGGCCTCTTTATCTCCATCATTGTAACTGCGGTGCTTGAGCTAAGATGGAGCGGTGTCAGCATTGAAGATTTATGGCGTA 2463
             |||| ||| ||| ||| ||| ||| ||| ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        258 AGTTTGTCTTGGCCTCTTTATCTCCATCATTGTAACTGCGGTGCTTGAGCTAAGATGGAGCGGTGTCAGCATTGAAGATTTATGGCGTA 337

Qy       2464 GTTTCTTGGCCTCTTTATCTCCATCATTGTAACTGCGGTGCTTGAGCTAAGATGGAGCGGTGTCAGCATTGAAGATTTATGGCGTA 2543
             || ||| ||| ||| ||| ||| ||| ||| ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        338 GTTTCTTGGCCTCTTTATCTCCATCATTGTAACTGCGGTGCTTGAGCTAAGATGGAGCGGTGTCAGCATTGAAGATTTATGGCGTA 417

Qy       2544 TGTCAGCATTGAAGATTTATGGCGTAA 2571
             ||||| ||| ||| ||| ||| |||
Db        418 TGTCAGCATTGAAGATTTATGGCGTAA 444


RESULT 3
BI132326
LOCUS       BI132326                 479 bp    mRNA    linear   EST 31-DEC-2001
DEFINITION  G135P19Y Populus cambium cDNA library Populus tremula x Populus
            tremuloides cDNA, mRNA sequence
ACCESSION   BI132326
VERSION     BI132326.1  GI:18016297
KEYWORDS    EST.
SOURCE      Populus tremula x Populus tremuloides
  ORGANISM  Populus tremula x Populus tremuloides
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids I; Malpighiales; Salicaceae; Populus.
REFERENCE   1  (bases 1 to 479)
  AUTHORS   Hertzberg,M., Aspeborg,H., Erlandsson,R., Bjorkbacka,H., Hiltonen
            ,T., Karlsson,J., Teeri,T., Gustafsson,P., Bahlerao,R., Jansson,S.,
            Nilsson,O., Sundberg,B., Nilsson,P., Uhlen,M., Sandberg,G. and
            Lundeberg,J.
  TITLE     Gene expression in Populus
  JOURNAL   Unpublished
COMMENT     Contact: Erlandsson R
            Department of Biotechnology
            Royal Institute of Technology
            Teknikringen 30, Stockholm S-10044, Sweden
            Tel: 46 8 790 8287
            Fax: 46 8 245452
            Email: rikerl@biochem.kth.se.
FEATURES            Location/Qualifiers
     source         1. .479
                    /organism="Populus tremula x Populus tremuloides"
                    /mol_type="mRNA"
                    /db_xref="taxon:47664"
                    /clone_lib="Populus cambium cDNA library"
                    /note="Organ: cambium"
BASE COUNT      111 a    116 c    113 g    139 t
ORIGIN

  Query Match             7.6%;  Score 245;  DB 12;  Length 479;
  Best Local Similarity   100.0%;  Pred. No. 7.8e-40;
  Matches  245;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy       2295 CAGACACTGTCCCCTCTGGTACGGGTTTGGAGGAGGCCGTCTTAAATGGCTCCAAAGGCT 2314
             ||  |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db        199 CAGACACTGTCCCCTCTGGTACGGGTTTGGAGGAGGCCGTCTTAAATGGCTCCAAAGGCT 258

Qy       2315 TGCGTATATAAACACCATTGTGTACCCATTTACATCCCTCGCTCTCATTGCCTATTGCAC 2374
             |||||||||||||||||||||  ||| |||||||||| ||| ||| ||||||||||||
Db        259 TGCGTATATAAACACCATTGTGTACCCATTTACATCCCTCGCTCTCATTGCCTATTGCAC 318

Qy       2375 AATTCCTGCAGTTTGTCTGCTCATCGGAAAATTCATCATATCAACGCTCTCAAACCTGGC 2434
             |||||| |||||| ||||| | |||| ||| ||| |||| |||| |||| ||||||||||
Db        319 AATTCCTGCAGTTTGTCTGCTCATCGGAAAATTCATCATATCAACGCTCTCAAACCTGGC 378

Qy       2435 AAGCATGCTGTTTCTTGGCCTCTTTATGTTCATCATTGTAACTTGCGGTGCTTGAGCTAAG 2494
             |||||||||||||| |||||||| |||| ||| ||| |||| |||| |||| ||||||||||
Db        379 AAGCATGCTGTTTCTTGGCCTCTTTATGTTCATCATTGTAACTTGCGGTGCTTGAGCTAAG 438

Qy       2495 ATGGA 2499
             |||||
Db        439 ATGGA 443


RESULT 4
BU895617
LOCUS       BU895617                571 bp    mRNA    linear   EST 17-OCT-2002
DEFINITION  X028D07 Populus wood cDNA library Populus tremula x Populus
            tremuloides cDNA 5 prime, mRNA sequence.
ACCESSION   BU895617
VERSION     BU895617.1  GI:24106824
KEYWORDS    EST.
SOURCE      Populus tremula x Populus tremuloides
  ORGANISM  Populus tremula x Populus tremuloides
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids I; Malpighiales; Salicaceae; Populus.
REFERENCE   1  (bases 1 to 571)
  AUTHORS   Unneberg,P., Bhalerao,R.P., Jansson,S. and Sterky,F.
  TITLE     The poplar tree transcriptome: Analysis of expressed sequence tags
            from multiple libraries
  JOURNAL   Unpublished
COMMENT     Contact: BHALERAO RUPALI P
            Umea Plant Science Center
            Department of Plant Physiology
            University of Umea, 901 87 Umea, Sweden
            Tel: +46 90 786 5279
            Fax: +46 90 786 6676
            Email: rupali.bhaleraoxxxxxxxxxx.xxx.xx
FEATURES             Location/Qualifiers
     source          1..571
                     /organism="Populus tremula x Populus tremuloides"
                     /mol_type="mRNA"
                     /db_xref="taxon:xxxx"
                     /tissue_type="wood"
                     /clone_lib="Populus xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx"
BASE COUNT      145 a    131 c    132 g    163 t
ORIGIN

  Query Match             7.1%;  Score 229;  DB 1;  Length 571;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-46;
  Matches  229;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2169 TTCAAAGGATCTGCACCCATCAAAGTTGTTCTATAGATTTTTAAGCTCCTCCGATGGGCT 2228
             |||||| |||||||||| |||| ||| |||||||| |||     | ||| |||||||||
Db        343 TTCAAAGGATCTGCACCCATCAAAGTTGTTCTATAGATTTTTAAGCTCCTCCGATGGGCT 402

Qy       2229 CTTGGTTCTGTGGAAATTTTCTTCAGCAGAAAATGCTGCTTTTGGTACGGGTTTGGAGGA 2288
             |||||| |||||||||||||||| ||| |||| |||| |||| |||| ||||||||||||
Db        403 CTTGGTTCTGTGGAAATTTTCTTCAGCAGAAAATGCTGCTTTTGGTACGGGTTTGGAGGA 462

Qy       2289 GGCCGTCTTAAATGGCTCCAAAGTTTGTTCTATAGATTTTTAAGCTCCTCCGATTTACA 2348
             |||||||||||||||||||||||| ||| |||| |||| |||| |||| |||||||||
Db        463 GGCCGTCTTAAATGGCTCCAAAGTTTGTTCTATAGATTTTTAAGCTCCTCCGATTTACA 522

Qy       2349 TCCCTCCCTCTCATTGCCTATTGCACAATTCCTGCAGTTTGTCTGCTCA 2397
             |||||||||||||||| |||||||||||||||||||||||||||||||
Db        523 TCCCTCCCTCTCATTGCCTATTGCACAATTCCTGCAGTTTGTCTGCTCA 571


RESULT 5
AI166604/c
LOCUS       AI166604                496 bp    mRNA    linear   EST 03-DEC-1998
DEFINITION  xylem.est.42 Poplar xylem Lambda ZAPII library Populus balsamifera
            subsp. trichocarpa cDNA 5', mRNA sequence.
ACCESSION   AI166604
VERSION     AI166604.1  GI:3857889
KEYWORDS    EST.
SOURCE      Populus balsamifera subsp. trichocarpa
  ORGANISM  Populus balsamifera subsp. trichocarpa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids I; Malpighiales; Salicaceae; Populus.
REFERENCE   1  (bases 1 to 496)
  AUTHORS   Sterky,F., Regan,S., Karlsson,J., Hertzberg,M., Rohde,A., Holmberg
            ,A., Amini,B., Bhalerao,R., Larsson,M., Villarroel,R., Van Montagu
            ,M., Sandberg,G., Olsson,O., Teeri,T.T., Boerjan,W., Gustafsson,P.,
            Uhlen,M., Sundberg,B. and Lundeberg,J.
  TITLE     Gene discovery in the wood-forming tissues of poplar: Analysis of
            5,692 expressed sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 95 (22), 13330-13335 (1998)
  MEDLINE   99007314
   PUBMED   9789088
COMMENT     Contact: Boerjan W
            Department of Genetics
            Flanders Interuniversity Institute for Biotechnology,University of
            Gent
            Ledeganckstraat 35, B-9000 Gent, Belgium
            Tel: +32 9 2645202
            Fax: +32 9 2645349
            Email: woboe@gengenp.rug.ac.be
            Seq primer: T3 primer
            High quality sequence stop: 496.
FEATURES             Location/Qualifiers
     source          1..496
                     /organism="Populus balsamifera subsp. trichocarpa"
                     /mol_type="mRNA"
                     /cultivar="Trichobel"
                     /db_xref="taxon:3694"
                     /tissue_type="Young, developing xylem"
                     /dev_stage="3-year-old actively growing tree (harvested in
                     May)"
                     /lab_host="E.coli"
                     /clone_lib="Poplar xylem Lambda ZAPII library"
                     /note="Vector: pBluescript SK; Site_1: EcoRI (5' end cDNA
                     ); Site_2: XhoI (3' end cDNA); Young developing xylem was
                     harvested from 3-year old, actively growing Populus
                     trichocarpa 'Trichobel' by peeling off the bark first and
                     then scraping on the remaining xylem. cDNA was prepared
                     and cloned directionally into Lambda ZAPII. Plasmid clones
                     of individual Lamdda clones were obtained by in vivo
                     excision."
BASE COUNT      166 a     95 c     88 g    147 t
ORIGIN

  Query Match             4.5%;  Score 146;  DB 9;  Length 496;
  Best Local Similarity   100.0%;  Pred. No. 3.7e-20;
  Matches  146;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2940 AATCCATTCGTTAACAAAGTTGATAACACCTTGGTTGCGGAGACCTGCATTTCCATTGAT 2999
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        374 AATCCATTCGTTAACAAAGTTGATAACACCTTGGTTGCGGAGACCTGCATTTCCATTGAT 315

Qy       3000 TGCTGAGCTACCTCCAATAAGTCTCTCCCAGTATTTTGGGGTTACAAAACCTTTGGGAAT 3059
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        314 TGCTGAGCTACCTCCAATAAGTCTCTCCCAGTATTTTGGGGTTACAAAACCTTTGGGAAT 255

Qy       3060 TGGAATATGATCCTCGTTGTAGTTTC 3085
             |||||||||||||||||||||||||
```

```
       254 TGGAATATGATCCTCGTTGTAGTTTC 229

ULT 6
97587
US        BU897587                 476 bp    mRNA    linear   EST 17-OCT-2002
INITION   X060B01 Populus wood cDNA library Populus tremula x Populus
          tremuloides cDNA 5 prime, mRNA sequence.
ESSION    BU897587
SION      BU897587.1  GI:24108794
WORDS     EST.
URCE      Populus tremula x Populus tremuloides
RGANISM   Populus tremula x Populus tremuloides
          Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
          Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
          ; eurosids I; Malpighiales; Salicaceae; Populus.
ERENCE    1  (bases 1 to 476)
AUTHORS   Unneberg,P., Bhalerao,R.R., Jansson,S. and Sterky,F.
TITLE     The poplar tree transcriptome: Analysis of expressed sequence tags
          from multiple libraries
JOURNAL   Unpublished
MMENT     Contact: BHALERAO RUPALI R.
          Umea Plant Science Center
          Department of Plant Physiology
          University of Umea, 901 87 Umea, Sweden
          Tel: +46 90 786 5279
          Fax: +46 90 786 6676
          Email: rupali.bhalerao@plantphys.umu.se.
ATURES              Location/Qualifiers
     source         1..476
                    /organism="Populus tremula x Populus tremuloides"
                    /mol_type="mRNA"
                    /db_xref="taxon:47664"
                    /tissue_type="wood"
                    /clone_lib="Populus wood cDNA library"
SE COUNT      114 a    108 c    103 g    151 t
IGIN

Query Match             4.2%;  Score 135;  DB 13;  Length 476;
Best Local Similarity   99.5%;  Pred. No. 6e-18;
Matches  185;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

       2738 TGCTGGATTCTCTGATGCACTCAACAAAGGATATGAAGCATGGGGGCCTCTCTTTGGCAA 2797
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
        182 TGCTGGATTCTCTGATGCACTCAACAAAGGATATGAAGCATGGGGGCCTCTCTTTGGCAA 241

       2798 GGTGTTCTTTGCTTTCTGGGTGATTCTTCATCTCTATCCATTCCTTAAAGGTCTAATGGG 2857
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
        242 GGTGTTCTTTGCTTTCTGGGTGATTCTTCATCTCTATCCATTCCTTAAAGGTCTAATGGG 301

       2858 TCGCCAAAACCTAACACCAACCATTGTTGTTCTCTGGTCAGTGCTGTTGGCCTCTGTCTT 2917
            ||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
        302 TCGCCAAAACCGAACACCAACCATTGTTGTTCTCTGGTCAGTGCTGTTGGCCTCTGTCTT 361

       2918 CTCTCT 2923
            ||||||
        362 CTCTCT 367


SULT 7
894292
CUS       BU894292                 648 bp    mRNA    linear   EST 17-OCT-2002
FINITION  X005H01 Populus wood cDNA library Populus tremula x Populus
          tremuloides cDNA 5 prime, mRNA sequence.
CCESSION  BU894292
RSION     BU894292.1  GI:24105357
YWORDS    EST.
URCE      Populus tremula x Populus tremuloides
RGANISM   Populus tremula x Populus tremuloides
          Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
          Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
          ; eurosids I; Malpighiales; Salicaceae; Populus.
FERENCE   1  (bases 1 to 648)
  AUTHORS   Unneberg,P., Bhalerao,R.R., Jansson,S. and Sterky,F.
  TITLE     The poplar tree transcriptome: Analysis of expressed sequence tags
            from multiple libraries
  JOURNAL   Unpublished
COMMENT     Contact: BHALERAO RUPALI R.
            Umea Plant Science Center
            Department of Plant Physiology
            University of Umea, 901 87 Umea, Sweden
            Tel: +46 90 786 5279
            Fax: +46 90 786 6676
            Email: rupali.bhalerao@plantphys.umu.se.
FEATURES             Location/Qualifiers
     source          1..685
                     /organism="Populus tremula x Populus tremuloides"
                     /mol_type="mRNA"
                     /db_xref="taxon:47664"
                     /tissue_type="wood"
BASE COUNT      [illegible]
ORIGIN

  Query Match             4.1%;  Score 135;  DB 13;  Length 648;
  Best Local Similarity   99.5%;  Pred. No. 4.4e-18;
  Matches  185;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2738 TGCTGGATTCTCTGATGCACTCAACAAAGGATATGAAGCATGGGGGCCTCTCTTTGGCAA 2797
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        116 TGCTGGATTCTCTGATGCACTCAACAAAGGATATGAAGCATGGGGGCCTCTCTTTGGCAA 175

Qy       2798 GGTGTTCTTTGCTTTCTGGGTGATTCTTCATCTCTATCCATTCCTTAAAGGTCTAATGGG 2857
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        176 GGTGTTCTTTGCTTTCTGGGTGATTCTTCATCTCTATCCATTCCTTAAAGGTCTAATGGG 235

Qy       2858 TCGCCAAAACCTAACACCAACCATTGTTGTTCTCTGGTCAGTGCTGTTGGCCTCTGTCTT 2917
             ||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db        236 TCGCCAAAACCGAACACCAACCATTGTTGTTCTCTGGTCAGTGCTGTTGGCCTCTGTCTT 295

Qy       2918 CTCTCT 2923
             ||||||
Db        296 CTCTCT 301


RESULT 8
BU897715
LOCUS       BU897715                 685 bp    mRNA    linear   EST 17-OCT-2002
DEFINITION  X067B10 Populus wood cDNA library Populus tremula x Populus
            tremuloides cDNA 5 prime, mRNA sequence.
ACCESSION   BU897715
VERSION     BU897715.1  GI:24108922
KEYWORDS    EST.
SOURCE      Populus tremula x Populus tremuloides
  ORGANISM  Populus tremula x Populus tremuloides
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids I; Malpighiales; Salicaceae; Populus.
REFERENCE   1  (bases 1 to 685)
  AUTHORS   Unneberg,P., Bhalerao,R.R., Jansson,S. and Sterky,F.
  TITLE     The poplar tree transcriptome: Analysis of expressed sequence tags
            from multiple libraries
  JOURNAL   Unpublished
COMMENT     Contact: BHALERAO RUPALI R.
            Umea Plant Science Center
            Department of Plant Physiology
            University of Umea, 901 87 Umea, Sweden
            Tel: +46 90 786 5279
            Fax: +46 90 786 6676
            Email: rupali.bhalerao@plantphys.umu.se.
FEATURES             Location/Qualifiers
     source          1..685
                     /organism="Populus tremula x Populus tremuloides"
                     /mol_type="mRNA"
                     /db_xref="taxon:47664"
                     /tissue_type="wood"
```

/clone_lib="Populus wood cDNA library"
BASE COUNT 165 a 142 c 142 g 233 t 1 others
ORIGIN

Query Match 4.2%; Score 135; DB 13; Length 683;
Best Local Similarity 99.5%; Pred. No. 4.4e-16;
Matches 185; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2738 TGCTGGATTCTCTGATGCACTCAACAAAGGATATGAAGCATGGGGCTCTCTTTGGCAA 2797
Db 53 TGCTGGATTCTCTGATGCACTCAACAAAGGATATGAAGCATGGGGCTCTCTTTGGCAA 112

Qy 2798 GGTGTTCTTTGCTTTCTGGGTGATTCTTCATCTCTATGCATCTTTAAAGGTCTAATGGG 2857
Db 113 GGTGTTCTTTGCTTTCTGGGTGATTCTTCATCTCTATGCATCTTTAAAGGTCTAATGGG 172

Qy 2858 TCGCCAAAACCTAACACCAACCATTGTTGTTCTCTGGTCATCTGCTGTTGGCCTCTGTCTT 2917
Db 173 TCGCCAAAACCGAACACCAACCATTGTTGTTCTCTGGTCATCTGCTGTTGGCCTCTGTCTT 232

Qy 2918 CTCTCT 2923
Db 233 CTCTCT 238

RESULT 9
BI128439
LOCUS BI128439 289 bp mRNA linear EST 31-DEC-2001
DEFINITION G076P02Y Populus cambium cDNA library Populus tremula x Populus tremuloides cDNA, mRNA sequence.
ACCESSION BI128439
VERSION BI128439.1 GI:18012416
KEYWORDS EST.
SOURCE Populus tremula x Populus tremuloides
ORGANISM Populus tremula x Populus tremuloides
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids ; eurosids I; Malpighiales; Salicaceae; Populus.
REFERENCE 1 (bases 1 to 289)
AUTHORS Hertzberg,M., Aspeborg,H., Erlandsson,R., Bjorkbacka,H., Hiltonen ,T., Karlsson,J., Teeri,T., Gustafsson,P., Bahlerao,R., Jansson,S., Nilsson,O., Sundberg,B., Nilsson,P., Uhlen,M., Sandberg,G. and Lundeberg,J.
TITLE Gene expression in Populus
JOURNAL Unpublished
COMMENT Contact: Erlandsson R
Department of Biotechnology
Royal Institute of Technology
Teknikringen 30, Stockholm S-10044, Sweden
Tel: 46 8 790 8287
Fax: 46 8 245452
Email: rikerl@biochem.kth.se.
FEATURES Location/Qualifiers
source 1..289
/organism="Populus tremula x Populus tremuloides"
/mol_type="mRNA"
/db_xref="taxon:47664"
/clone_lib="Populus cambium cDNA library"
/note="Organ: cambium"
BASE COUNT 72 a 65 c 63 g 91 t
ORIGIN

Query Match 4.1%; Score 131; DB 13; Length 289;
Best Local Similarity 100.0%; Pred. No. 5.6e-17;
Matches 131; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2738 TGCTGGATTCTCTGATGCACTCAACAAAGGATATGAAGCATGGGGCTCTCTTTGGCAA 2797
Db 122 TGCTGGATTCTCTGATGCACTCAACAAAGGATATGAAGCATGGGGCTCTCTTTGGCAA 181

Qy 2798 GGTGTTCTTTGCTTTCTGGGTGATTCTTCATCTCTATGCATCTTTAAAGGTCTAATGGG 2857
Db 182 GGTGTTCTTTGCTTTCTGGGTGATTCTTCATCTCTATGCATCTTTAAAGGTCTAATGGG 241

Qy 2858 TCGCCAAAACC 2868
Db 242 TCGCCAAAACC 252

RESULT 10
BI129538
LOCUS BI129538 287 bp mRNA linear EST 31-DEC-2001
DEFINITION G092P15Y Populus cambium cDNA library Populus tremula x Populus tremuloides cDNA, mRNA sequence.
ACCESSION BI129538
VERSION BI129538.1 GI:18013509
KEYWORDS EST.
SOURCE Populus tremula x Populus tremuloides
ORGANISM Populus tremula x Populus tremuloides
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids ; eurosids I; Malpighiales; Salicaceae; Populus.
REFERENCE 1 (bases 1 to 287)
AUTHORS Hertzberg,M., Aspeborg,H., Erlandsson,R., Bjorkbacka,H., Hiltonen ,T., Karlsson,J., Teeri,T., Gustafsson,P., Bahlerao,R., Jansson,S., Nilsson,O., Sundberg,B., Nilsson,P., Uhlen,M., Sandberg,G. and Lundeberg,J.
TITLE Gene expression in Populus
JOURNAL Unpublished
COMMENT Contact: Erlandsson R
Department of Biotechnology
Royal Institute of Technology
Teknikringen 30, Stockholm S-10044, Sweden
Tel: 46 8 790 8287
Fax: 46 8 245452
Email: rikerl@biochem.kth.se.
FEATURES Location/Qualifiers
source 1..287
/organism="Populus tremula x Populus tremuloides"
/mol_type="mRNA"
/db_xref="taxon:47664"
/clone_lib="Populus cambium cDNA library"
/note="Organ: cambium"
BASE COUNT 66 a 78 c 64 g 79 t
ORIGIN

Query Match 3.8%; Score 123; DB 12; Length 287;
Best Local Similarity 100.0%; Pred. No. 2.2e-15;
Matches 123; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2247 TTCTTTAGCAGACACTGTCCCCTCTGGTACGGGTTTGGAGGAGGCCGTCTTAAATGGCTC 2306
Db 152 TTCTTTAGCAGACACTGTCCCCTCTGGTACGGGTTTGGAGGAGGCCGTCTTAAATGGCTC 211

Qy 2307 CAAAGGCTTGCGTATATAAACACCATTGTGTACCCATTTACATCCCTCCCTCTCATTGCC 2366
Db 212 CAAAGGCTTGCGTATATAAACACCATTGTGTACCCATTTACATCCCTCCCTCTCATTGCC 271

Qy 2367 TAT 2369
Db 272 TAT 274

RESULT 11
BU894130
LOCUS BU894130 434 bp mRNA linear EST 17-OCT-2002
DEFINITION X003C01 Populus wood cDNA library Populus tremula x Populus tremuloides cDNA 5 prime, mRNA sequence.
ACCESSION BU894130
VERSION BU894130.1 GI:24105195
KEYWORDS EST.
SOURCE Populus tremula x Populus tremuloides
ORGANISM Populus tremula x Populus tremuloides
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids ; eurosids I; Malpighiales; Salicaceae; Populus.

```
ERENCE    1  (bases 1 to 434)
UTHORS    Unneberg,P., Bhalerao,R.R., Jansson,S. and Sterky,F.
ITLE      The poplar tree transcriptome: Analysis of expressed sequence tags
          from multiple libraries
OURNAL    Unpublished
MENT      Contact: BHALERAO RUPALI R.
          Umea Plant Science Center
          Department of Plant Physiology
          University of Umea, 901 87 Umea, Sweden
          Tel: +46 90 786 5279
          Fax: +46 90 786 6676
          Email: rupali.bhalerao@plantphys.umu.se.
ATURES             Location/Qualifiers
  source           1..434
                   /organism="Populus tremula x Populus tremuloides"
                   /mol_type="mRNA"
                   /db_xref="taxon:47664"
                   /tissue_type="wood"
                   /clone_lib="Populus wood cDNA library"
SE COUNT      132 a     66 c    125 g    111 t
GIN

Query Match             3.5%;  Score 114;  DB 13;  Length 434;
Best Local Similarity   100.0%;  Pred. No. 9.7e-14;
Matches  114;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

        31 TGAAGTTGTAAAGAGCTGGTAAAGTGGTAATAAGCAAGATGATGGAATCTGGGGCTCCTA 90
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
        24 TGAAGTTGTAAAGAGCTGGTAAAGTGGTAATAAGCAAGATGATGGAATCTGGGGCTCCTA 83

        91 TATGCCATACCTGTGGTGAACAGGTGGGGCATGATGCAAATGGGGAGCTATTTG 144
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
        84 TATGCCATACCTGTGGTGAACAGGTGGGGCATGATGCAAATGGGGAGCTATTTG 137


SULT 12
127228
CUS       BI127228                435 bp    mRNA    linear   EST 31-DEC-2001
FINITION  G057P27Y Populus cambium cDNA library Populus tremula x Populus
          tremuloides cDNA, mRNA sequence.
CESSION   BI127228
RSION     BI127228.1  GI:18011199
YWORDS    EST.
URCE      Populus tremula x Populus tremuloides
ORGANISM  Populus tremula x Populus tremuloides
          Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
          Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
          ; eurosids I; Malpighiales; Salicaceae; Populus.
FERENCE   1  (bases 1 to 435)
AUTHORS   Hertzberg,M., Aspeborg,H., Erlandsson,R., Bjorkbacka,H., Hiltonen
          ,T., Karlsson,J., Teeri,T., Gustafsson,P., Bahlerao,R., Jansson,S.,
          Nilsson,O., Sundberg,B., Nilsson,P., Uhlen,M., Sandberg,G. and
          Lundeberg,J.
TITLE     Gene expression in Populus
JOURNAL   Unpublished
MMENT     Contact: Erlandsson R
          Department of Biotechnology
          Royal Institute of Technology
          Teknikringen 30, Stockholm S-10044, Sweden
          Tel: 46 8 790 8287
          Fax: 46 8 245452
          Email: rikerl@biochem.kth.se.
ATURES             Location/Qualifiers
  source           1..435
                   /organism="Populus tremula x Populus tremuloides"
                   /mol_type="mRNA"
                   /db_xref="taxon:47664"
                   /clone_lib="Populus cambium cDNA library"
                   /note="Organ: cambium"
SE COUNT      114 a    110 c     98 g    113 t
IGIN

Query Match             3.0%;  Score 98;  DB 12;  Length 435;
```

```
Best Local Similarity   [illegible]%;  Pred. No. [illegible]e-10;
Matches  148;  Conservative  [illegible];  Mismatches    [illegible];  Indels    0;  Gaps    0;

Qy       2082 GGGTCAGTCACT[illegible]ATATCTTAA[illegible]AAGATGCACTGCCGAGGATGGAGAT[illegible]A 2141
Db         91 GGGTCAGTCACT[illegible]ATATCTTAA[illegible]AAGATGCACTGCCGAGGATGGAGAT[illegible]A 150

Qy       2142 ATTTACT[illegible]ATCTGCACCCATCAACCTGTCTGAT 2201
Db        151 ATTTACT[illegible]AAAGGATCTGCACCCATCAACCTGTCTGAT 210

Qy       2202 AGATTGCACTAG[illegible]
Db        211 AGATTGCAC[illegible]


RESULT 13
BI129892
LOCUS       BI129892    [illegible]    mRNA    linear   EST 31-DEC-2001
DEFINITION  [illegible] Populus tremula x Populus
            tremuloides cDNA, mRNA sequence.
ACCESSION   BI129892
VERSION     BI129892.1  GI:[illegible]
KEYWORDS    EST.
SOURCE      Populus tremula x Populus tremuloides
  ORGANISM  Populus tremula x Populus tremuloides
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids I; Malpighiales; Salicaceae; Populus.
REFERENCE   1  (bases 1 to 126)
  AUTHORS   Hertzberg,M., Aspeborg,H., Erlandsson,R., Bjorkbacka,H., Hiltonen
            ,T., Karlsson,J., Teeri,T., Gustafsson,P., Bahlerao,R., Jansson,S.,
            Nilsson,O., Sundberg,B., Nilsson,P., Uhlen,M., Sandberg,G. and
            Lundeberg,J.
  TITLE     Gene expression in Populus
  JOURNAL   Unpublished
COMMENT     Contact: Erlandsson R
            Department of Biotechnology
            Royal Institute of Technology
            Teknikringen 30, Stockholm S-10044, Sweden
            Tel: 46 8 790 [illegible]
            Fax: 46 8 245452
            Email: rikerl@biochem.kth.se.
FEATURES             Location/Qualifiers
     source          1..126
                     /organism="Populus tremula x Populus tremuloides"
                     /mol_type="mRNA"
                     /db_xref="taxon:47664"
                     /clone_lib="Populus cambium cDNA library"
                     /note="Organ: cambium"
BASE COUNT       36 a     24 [illegible]     28 g     3[illegible] t
ORIGIN

  Query Match             [illegible]%;  Score [illegible];  DB 12;  Length 126;
  Best Local Similarity   [illegible]%;  Pred. No. [illegible]e-09;
  Matches   93;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1451 TTGTGATCA[illegible]TAT[illegible]AAACAATAGCAAG[illegible]T[illegible]TTCGAGAGGCAATGTGCAT[illegible]TGATGGA 1510
Db         34 TTGTGATCA[illegible]TAT[illegible]AAACAATAGCAAG[illegible]T[illegible]TTCGAGAGGCAATGTGCAT[illegible]TGATGGA 93

Qy       1511 CCCACAAGTAGGT[illegible]AGATGTATGCTATGTG[illegible]A 1543
Db         94 CCCACAAGTAGGT[illegible]AGATGTATGCTATGTG[illegible]A 126


RESULT 14
BU897069
LOCUS       BU897069                337 bp    mRNA    linear   EST 17-OCT-2002
DEFINITION  X050E11 Populus wood cDNA library Populus tremula x Populus
            tremuloides cDNA 5 prime, mRNA sequence.
ACCESSION   BU897069
VERSION     BU897069.1  GI:24108276
```

```
KEYWORDS    EST.
SOURCE      Populus tremula x Populus tremuloides
  ORGANISM  Populus tremula x Populus tremuloides
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids I; Malpighiales; Salicaceae; Populus.
REFERENCE   1  (bases 1 to 337)
  AUTHORS   Unneberg,P., Bhalerao,R.R., Jansson,S. and Sterky,F.
  TITLE     The poplar tree transcriptome: Analysis of expressed sequence tags
            from multiple libraries
  JOURNAL   Unpublished
COMMENT     Contact: BHALERAO RUPALI R
            Umea Plant Science Center
            Department of Plant Physiology
            University of Umea, 901 87 Umea, Sweden
            Tel: +46 90 786 5279
            Fax: +46 90 786 6676
            Email: rupali.bhaleraoaplantphysiol.umu.se.
FEATURES             Location/Qualifiers
     source          1..337
                     /organism="Populus tremula x Populus tremuloides"
                     /mol_type="mRNA"
                     /db_xref="taxon:47664"
                     /tissue_type="wood"
                     /clone_lib="Populus wood cDNA library"
BASE COUNT       84 a     80 c     92 g     81 t
ORIGIN

  Query Match             2.5%;  Score 82;  DB 13;  Length 337;
  Best Local Similarity   100.0%;  Pred. No. 2.8e-07;
  Matches   82;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2340 AATCCATTCGTTAACAAAGTTGATAAAACCTTGGTTGCAAAAACCTGCATTTCCATTGAT 2999
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         97 AATCCATTCGTTAACAAAGTTGATAAAACCTTGGTTGCAAAAACCTGCATTTCCATTGAT 156

Qy       3000 TGCTGAGCTACCTCCAATAAGT 3021
             |||||||||||||||||||||||
Db        157 TGCTGAGCTACCTCCAATAAGT 178


RESULT 15
BU896785
LOCUS       BU896785                 233 bp    mRNA    linear   EST 17-OCT-2002
DEFINITION  X045F11 Populus wood cDNA library Populus tremula x Populus
            tremuloides cDNA 5 prime, mRNA sequence.
ACCESSION   BU896785
VERSION     BU896785.1  GI:24107992
KEYWORDS    EST.
SOURCE      Populus tremula x Populus tremuloides
  ORGANISM  Populus tremula x Populus tremuloides
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids I; Malpighiales; Salicaceae; Populus.
REFERENCE   1  (bases 1 to 233)
  AUTHORS   Unneberg,P., Bhalerao,R.R., Jansson,S. and Sterky,F.
  TITLE     The poplar tree transcriptome: Analysis of expressed sequence tags
            from multiple libraries
  JOURNAL   Unpublished
COMMENT     Contact: BHALERAO RUPALI R.
            Umea Plant Science Center
            Department of Plant Physiology
            University of Umea, 901 87 Umea, Sweden
            Tel: +46 90 786 5279
            Fax: +46 90 786 6676
            Email: rupali.bhaleraoaplantphysiol.umu.se.
FEATURES             Location/Qualifiers
     source          1..233
                     /organism="Populus tremula x Populus tremuloides"
                     /mol_type="mRNA"
                     /db_xref="taxon:47664"
                     /tissue_type="wood"
                     /clone_lib="Populus wood cDNA library"
BASE COUNT       83 a     43 c     61 g     46 t
ORIGIN

  Query Match             2.4%;  Score 77;  DB 13;  Length 233;
  Best Local Similarity   100.0%;  Pred. No. 3.7e-06;
  Matches   77;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        498 GAACAGCAGATGGAAGAGAAACCGTCTGCAGAGGCTTCGGAGCCGCTTTCAATTGTTTAT 557
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        147 GAACAGCAGATGGAAGAGAAACCGTCTGCAGAGGCTTCGGAGCCGCTTTCAATTGTTTAT 206

Qy        558 CCAATTCCACGCAACAA 574
              |||||||||||||||||
Db        207 CCAATTCCACGCAACAA 223


Search completed: October 31, 2003, 09:25:37
Job time : 6350.45 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 29, 2003, 18:25:23 ; Search Time 20.8085 Seconds
(without alignments)
2210.255 Million cell updates/sec

Title: US-09-980-043-2
Perfect score: 5240
Sequence: 1 MMESGAPICHTCGEQVGHDA.....PFVNKVSNTLVAETCISIDC 978

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	230	4.4	874	1	BCSA_SALTI	Q8z291 salmonella
2	228	4.4	874	1	BCSA_SALTY	Q93in2 salmonella
3	223	4.3	754	1	BCA1_ACEXY	P19449 acetobacter
4	222	4.2	872	1	BCSA_ECO57	Q8x517 escherichia
5	221	4.2	872	1	BCSA_ECOLI	P37653 escherichia
6	217	4.1	756	1	BCA2_ACEXY	O82859 acetobacter
7	214	4.1	739	1	BCSA_PSEFL	P58931 pseudomonas
8	211.5	4.0	1518	1	BCA4_ACEXY	Q9rbj2 acetobacter
9	211.5	4.0	1518	1	BCA5_ACEXY	Q9wx75 acetobacter
10	208	4.0	745	1	BCA3_ACEXY	Q9wx61 acetobacter
11	207.5	4.0	1596	1	ACS2_ACEXY	Q59167 acetobacter
12	181	3.5	729	1	BCSA_XANAC	P58932 xanthomonas
13	174.5	3.3	1550	1	ACS1_ACEXY	P21877 acetobacter
14	127.5	2.4	1177	1	A11B_HUMAN	Q9y2g3 homo sapien
15	117	2.2	1402	1	ATCX_SCHPO	Q09891 schizosacch
16	113.5	2.2	959	1	LONM_HUMAN	P36776 homo sapien
17	111.5	2.1	1433	1	DPO3_BACHD	Q9ka72 bacillus ha
18	111	2.1	752	1	HEPA_HSV2H	P89431 herpes simp
19	111	2.1	1169	1	A11B_RABIT	Q9n0z4 oryctolagus
20	109	2.1	1369	1	MOKE_SCHPO	Q9y7c4 schizosacch
21	108.5	2.1	1133	1	LONM_YEAST	P36775 saccharomyc
22	108.5	2.1	3329	1	BRC2_MOUSE	P97929 mus musculu
23	108.5	2.1	3712	1	ACVS_CEPAC	P25464 cephalospor
24	107	2.0	5085	1	PCLO_RAT	Q9jks6 rattus norv
25	106	2.0	1065	1	KDGI_HUMAN	O75912 homo sapien
26	105	2.0	483	1	SYE_BACSU	P22250 bacillus su
27	104.5	2.0	1103	1	A11C_HUMAN	Q8rb49 homo sapien
28	104.5	2.0	2515	1	TUD_DROME	P25823 drosophila
29	104	2.0	454	1	TR1A_MOUSE	P25118 mus musculu
30	103.5	2.0	554	1	DCS2_GOSAR	Q39760 gossypium a
31	103.5	2.0	6632	1	UN89_CAEEL	O01761 caenorhabdi
32	102	1.9	400	1	BMR2_BACSU	P39843 bacillus su
33	101.5	1.9	889	1	RPA1_METVA	P41556 methanococc
34	100.5	1.9	383	1	COS7_YEAST	Q07788 saccharomyc
35	100.5	1.9	424	1	CBPT_THEVU	P29068 thermoactin
36	100.5	1.9	441	1	YCDQ_ECOLI	P75905 escherichia
37	100.5	1.9	1736	1	ZO1_HUMAN	Q07157 homo sapien
38	99.5	1.9	347	1	A11C_MOUSE	Q9qzw0 mus musculu
39	99.5	1.9	696	1	YO12_CAEEL	P34668 caenorhabdi
40	99.5	1.9	805	1	SEC6_YEAST	P32844 saccharomyc
41	99.5	1.9	814	1	TOP1_SCHPO	P07799 schizosacch
42	99.5	1.9	1019	1	TOP1_USTMA	P41511 ustilago ma
43	99.5	1.9	5038	1	PCLO_MOUSE	Q9qyx7 mus musculu
44	99	1.9	536	1	COX1_ALLMA	P80440 allomyces m
45	99	1.9	796	1	ABAA_EMENI	P20945 emericella

ALIGNMENTS

```
RESULT 1
BCSA_SALTI
ID   BCSA_SALTI     STANDARD;      PRT;   874 AA.
AC   Q8Z291;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Cellulose synthase catalytic subunit [UDP-forming] (EC 2.4.1.12).
GN   BCSA OR STY4181 OR T3898.
OS   Salmonella typhi.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Salmonella.
OX   NCBI_TaxID=601;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CT18;
RX   MEDLINE=21534947; PubMed=11677608;
RA   Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA   Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
RA   Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,
RA   Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA   Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA   Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA   Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA   Whitehead S., Barrell B.G.;
RT   "Complete genome sequence of a multiple drug resistant Salmonella
RT   enterica serovar Typhi CT18.";
RL   Nature 413:848-852(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Ty2 / ATCC 700931;
RX   MEDLINE=22531367; PubMed=12644504;
RA   Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA   Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;
RT   "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
RT   and CT18.";
RL   J. Bacteriol. 185:2330-2337(2003).
CC   -!- FUNCTION: Catalytic subunit of cellulose synthase. It polymerizes
CC       uridine 5'-diphosphate glucose to cellulose, which is produced as
CC       an extracellular component for mechanical and chemical protection
CC       at the onset of the stationary phase, when the cells exhibit
CC       multicellular behavior (rdar morphotype). Co-expression of
CC       cellulose and thin aggregative fimbriae leads to a hydrophobic
CC       network with tightly packed cells embedded in a highly inert
CC       matrix (By similarity).
CC   -!- CATALYTIC ACTIVITY: UDP-glucose + {(1,4)-beta-D-glucosyl}(N) = UDP
CC       + {(1,4)-beta-D-glucosyl}(N+1).
CC   -!- COFACTOR: Magnesium (By similarity).
CC   -!- ENZYME REGULATION: Activated by bis-(3'-5') cyclic diguanylic acid
CC       (c-di-GMP) (By similarity).
CC   -!- PATHWAY: Bacterial cellulose biosynthesis.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC       (Potential).
CC   -!- DOMAIN: There are two conserved domains in the globular part of
CC       the protein: the N-terminal domain (domain A) contains
CC       the conserved DXD motif and is possibly involved in catalysis and
```

substrate binding. The C-terminal domain (domain B) contains the QXXRW motif and is present only in processive glycosyl transferases. It could be involved in the processivity function of the enzyme, possibly required for holding the growing glycan chain in the active site.
-!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).

EMBL; AL627281; CAD08006.1; -.
EMBL; AE016847; AAO71373.1; -.
InterPro; IPR003919; Cellsynth_A.
InterPro; IPR001173; Glyco_trans_2.
Pfam; PF00535; Glycos_transf_2; 1.
PRINTS; PR01439; CELLSNTHASEA.
Cellulose biosynthesis; Transferase; Glycosyltransferase; Transmembrane; Inner membrane; Complete proteome.

DOMAIN	271	364	CATALYTIC SUBDOMAIN A.
DOMAIN	441	501	CATALYTIC SUBDOMAIN B.
TRANSMEM	30	50	POTENTIAL.
TRANSMEM	151	171	POTENTIAL.
TRANSMEM	173	193	POTENTIAL.
TRANSMEM	230	250	POTENTIAL.
TRANSMEM	525	545	POTENTIAL.
TRANSMEM	547	567	POTENTIAL.
TRANSMEM	592	612	POTENTIAL.
TRANSMEM	634	654	POTENTIAL.
TRANSMEM	668	688	POTENTIAL.
TRANSMEM	833	853	POTENTIAL.
ACT_SITE	313	313	POTENTIAL.
ACT_SITE	457	457	POTENTIAL.
BINDING	360	360	SUBSTRATE (POTENTIAL).
BINDING	362	362	SUBSTRATE (POTENTIAL).
SEQUENCE	874 AA;	100020 MW;	3395D09CDCF51990 CRC64;

Query Match 4.4%; Score 230; DB 1; Length 874;
Best Local Similarity 18.9%; Pred. No. 5e-09;
Matches 153; Conservative 81; Mismatches 217; Indels 358; Gaps 31;

```
Qy      176 VIIMRLVILGLFFHFRIT------NPVDSAFGLWLTSVICEIWFAFSWVLDQFPKWNPVN 229
            :|:: | :   :  :| |       :||    || |       |     ||  |    |:|
Db      202 LIVLSLTVSCRYIWWRYTSTLNWDDPVSLVCGLILLFAETYAWIVL--VLGYFQVVWPLN 259

Qy      230 RETYIERLSARYEREGEPSQLAGVDFFVSTVDPLKEPPLITANTVLSILAVDYPVDKVSC 289
            |:              | ||   || || | :   |   :  ||: : | :|:| ||::
Db      260 RQP--------VPLPKEMSQWPTVDIFVPTYN---EDLNVVKNTIYASLGIDWPKDKLNI 308

Qy      290 YVSDDGAAMLSFESLVETAEFARKWVPFCKKFSIEPRAPEFYFSQKIDYLKDKVQPSFVK 349
            :: |||      ||     :|||
Db      309 WILDDGGR----ESF---RQFAR---------------------------------- 324

Qy      350 ERRAMKRDYEEYKVRVNALVAKAQKTPEEGWTMQDGTPWPGNNTRDHPGHDSGLPWEILG 409
                                                          | |
Db      325 ----------------------------------------------HVG---------- 327

Qy      410 ARDIEGNELPRLVYVSREKRPGYQHHKKAGAENALVRVSAVLTNAPYILNVDCDHYVNNS 469
                    : |::|     ::| |     |||           ::   ||||    |
Db      328 -----------VHYIARAT---HEHAKAGNINNALKH-----AKGEFVAIFDCDHVPTRS 368

Qy      470 KAVREAMCILMDPQVGRDVCYVQFPQRFDGIDKSDR-------YANRNVVFFDVNMKGLD 522
                  | : |    :  :| | |  |  | :|       |   :|: :   | |
Db      369 FLQMTMGWFLKEKQ----LAMMQTPHHFFSPDPFERNLGRFRKTPNEGTLFYGLVQDGND 424

Qy      523 GIQGPVYVGTGCVFNRQALYGYGPPSMPSLRKRKDSSSCFSCCCPSKKKPAQDPAEVYRD 582
                : |:                                  |   ::||
Db      425 MWDATFFCGS------------------------------CAVIRRKP---------- 442
```

```
Qy      583 AKREDLNAAIFNLTF[illegible]ENYDEHERSMLISQLSFFKTFGLSSVFIESTLMENGGVPESANS 642
                                                                  | | |::
Db      443 -------  -----      --  -  ---------------------LDEIGGI ----- 449

Qy      643 PPFIKEAIQVIGCC[illegible]FKTEWGKQIGWIYGSVTEDILSGFKMHCRGWRSIYCMPVRPAFK 702
                  :::                          :||:|  :  ::| ||: | |  . : |
Db      450 ----- AVE- ---          -  -  -- TVTEDAHTSLRLHRRGYTSAY-MRI-PQSA 480

Qy      703 GSAPINLSDRLHQVL[illegible]WALGSVEIFFSRHCPLWYGFGGGRLKWLQRLAYINTIVYPFTSL 762
            | |  :|:  :   :|   | |: |   : ||   :| | ||  |||: |:| : :  : :
Db      481 GLATESLSAHIGGRF[illegible]WARGMVQIFRLDN-PL---FGKG-LKLAQRLCYLNAMFHFLSGI 535

Qy      763 PLIAYCTIPAVCLLT[illegible]KFIIPTLSNLASMLFIALFISIIVTAVLELRWSGVSIEDLWRNE 822
            | : :  |   :    :|       |  : : |.  ::  ::    :   |:  :|:
Db      536 PRLIFLTAPLAFLIF[illegible]AYII- ---YAPALMIALF---VIPHMVHASLTNSKIQGKYRH- 586

Qy      823 QFWVIGGVSAHLFAV[illegible]QQFLKMLAGIDTNFTVTAKAAEDAEFGELYMVKWTTLLIPPTTL 882
              ||                                                |:|       : :|  :
Db      587 SFW  -                -  -  -----  ---------SEIYETVLAWYIAPPTLV 607

Qy      883 LIINMSGVAGFSDAI[illegible]----YFAW - [illegible]LPGKVFFAFWVILHLYPFLKGL----M 929
              ::|     :  :|     :  |     :       : | |:|:|     |:
Db      608 ALINPH  KGKFNL[illegible]YWVIGKI------YIFLVLLNILGVAAGVWPYYY 659

Qy      930 GRQNLTPTIVVLWS[illegible]LASVFSLVWVKIN 966
            | :| | |::|          ||||.  |
Db      660 GPENETLTVIV--    -  --SLVWVFYN 678
```

RESULT 2
BCSA_SALTY
ID BCSA_SALTY STANDARD; PRT; 874 AA.
AC Q93IN2;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cellulose synthase catalytic subunit [UDP forming] (EC 2.4.1.12).
GN BCSA OR STM3619.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 14028;
RX MEDLINE=21160181; PubMed=11260463;
RA Zogaj X., Nimtz M., Rohde M., Bokranz W., Roemling U.;
RT "The multicellular morphotypes of Salmonella typhimurium and
RT Escherichia coli produce cellulose as the second component of the
RT extracellular matrix.";
RL Mol. Microbiol. 39:1452-1463(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2;
RX MEDLINE=21927388; PubMed=11929533;
RA Solano C., Garcia B., Valle J., Berasain C., Ghigo J.-M., Gamazo C.,
RA Lasa I.;
RT "Genetic analysis of Salmonella enteritidis biofilm formation:
RT critical role for cellulose.";
RL Mol. Microbiol. 43:793-808(2002).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.";
RL Nature 413:852-856(2001).
CC -!- FUNCTION: Catalytic subunit of cellulose synthase. It polymerizes

```
CC       uridine 5'-diphosphate glucose to cellulose, which is produced as
CC       an extracellular component for mechanical and chemical protection
CC       at the onset of the stationary phase, when the cells exhibit
CC       multicellular behavior (rdar morphotype). Co-expression of
CC       cellulose and thin aggregative fimbriae leads to a hydrophobic
CC       network with tightly packed cells embedded in a highly inert
CC       matrix.
CC   -!- CATALYTIC ACTIVITY: UDP-glucose + {(1,4)-beta-D-glucosyl}(N) = UDP
CC       + {(1,4)-beta-D-glucosyl}(N+1).
CC   -!- COFACTOR: Magnesium (By similarity).
CC   -!- ENZYME REGULATION: Activated by bis-(3'-5') cyclic dimanylic acid
CC       (c-di GMP) (By similarity).
CC    !  PATHWAY: Bacterial cellulose biosynthesis.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC       (Potential).
CC   -!- DOMAIN: There are two conserved domains in the globular part of
CC       the protein: the N-terminal domain (domain A) contains
CC       the conserved DXD motif and is possibly involved in catalysis and
CC       substrate binding. The C-terminal domain (domain B) contains the
CC       QXXRW motif and is present only in processive glycosyl
CC       transferases. It could be involved in the processivity function of
CC       the enzyme, possibly required for holding the growing glycan chain
CC       in the active site.
CC    !- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE 2 FAMILY 2.
CC   ---- ------- ------------------- ------------ ---------- ----------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --- ------------------------ - -- ---------- ---
DR   EMBL; AJ315270; CAC44015.1;
DR   EMBL; AE008867; AAL22479.1; -.
DR   EMBL; AJ315148; CAC86199.1;
DR   StyGene; SG?????; bcsA.
DR   InterPro; IPR003919; Cellsynth_A
DR   InterPro; IPR005150; Cellulose_synt
DR   InterPro; IPR001173; Glyco_trans_2.
DR   Pfam; PF03552; Cellulose_synt; 1
DR   Pfam; PF00535; Glycos_transf_2; 1
DR   PRINTS; PR01439; CELLSNTHASEA.
KW   Cellulose biosynthesis; Transferase; Glycosyltransferase;
KW   Transmembrane; Inner membrane; Complete proteome.
FT   DOMAIN       271     364       CATALYTIC SUBDOMAIN A.
FT   DOMAIN       441     501       CATALYTIC SUBDOMAIN
FT   TRANSMEM      30      50       POTENTIAL.
FT   TRANSMEM     161     171       POTENTIAL.
FT   TRANSMEM     173     193       POTENTIAL.
FT   TRANSMEM     230     250       POTENTIAL.
FT   TRANSMEM     525     545       POTENTIAL.
FT   TRANSMEM     547     567       POTENTIAL.
FT   TRANSMEM     592     612       POTENTIAL.
FT   TRANSMEM     634     654       POTENTIAL.
FT   TRANSMEM     668     688       POTENTIAL.
FT   TRANSMEM     833     853       POTENTIAL.
FT   ACT_SITE     313     313       POTENTIAL.
FT   ACT_SITE     457     457       POTENTIAL.
FT   BINDING      360     360       SUBSTRATE (POTENTIAL).
FT   BINDING      362     362       SUBSTRATE (POTENTIAL).
SQ   SEQUENCE   874 AA;  100043 MW;  4C9421B8 8606310A CRC64;

  Query Match             4.4%;  Score 228;  DB 1;  Length 874;
  Best Local Similarity   18.9%;  Pred. No. 7.1e-09;
  Matches  153;  Conservative   80;  Mismatches  218;  Indels  358;  Gaps   31;

Qy        176 VIIMRLVILGLFFHFRIT- -   NPVDSAFGLWLTSVIF WFAFSWVLDQFPKWNPVN 229
              :|:: | :   :  :| :       :||    || :      |     ||  |    |:|
Db        202 LIVLSLTVSCRYIWWRYTSTLNWI DPVSLV GLILLFAL AWIVL -VLGYFQVVWPLN 259

Qy        230 RETYIERLSARYEREGEPSQLAGVDFFVSTVDPLKEPPLI ANTVLSILAVDYPVDKVSC 289
              |:             | |      || : :  '  . ||:  | :|:| .|::
```

```
Db        260 RQP--------VPLPKEMSQWPTVDIFVPTYN---EDLNVVKNTIYASLGIDWPKDKLNI 308

Qy        290 YVSDDGAAMLSFESLVETAEFARKWVPFCKKFSIEPRAPEFYFSQKIDYLKDKVQPSFVK 349
              :: |||      ||      |||
Db        309 WILDDGGR----ESF---RHFAR---------------------------------- 324

Qy        350 ERRAMKRDYEEYKVRVNALVAKAQKTPEEGWTMQDGTPWPGNNTRDHPGHDSGLPWEILG 409
                                                                | |
Db        325 ----------------------------------------------HVG---------- 327

Qy        410 ARDIEGNELPRLVYVSREKRPGYQHHKKAGAENALVRVSAVLTNAPYILNVDCDHYVNNS 469
                        : |::|     ::| |    |||          ::    ||||    |
Db        328 -----------VHYIARTT---HEHAKAGNINNALKH-----AKGEFVAIFDCDHVPTRS 368

Qy        470 KAVREAMCILMDPQVGRDVCYVQFPQRFDGIDKSDR-------YANRNVVFFDVNMKGLD 522
                    | : |    :  :| | |   |  :|          |   :|: :   | |
Db        369 FLQMTMGWFLKEKQ----LAMMQTPHHFFSPDPFERNLGRFRKTPNEGTLFYGLVQDGND 424

Qy        523 GIQGPVYVGTGCVFNRQALYGYGPPSMPSLRKRKDSSSCFSCCCPSKKKPAQDPAEVYRD 582
                  : |:                                    |   ::||
Db        425 MWDATFFCGS-------------------------------CAVIRRKP---------- 442

Qy        583 AKREDLNAAIFNLTEIDNYDEHERSMLISQLSFEKTFGLSSVFIESTLMENGGVPESANS 642
                                                                | | ||:
Db        443 ----------------------------------------------LDEIGGI------ 449

Qy        643 PPFIKEAIQVIGCGYEEKTEWGKQIGWIYGSVTEDILSGFKMHCRGWRSIYCMPVRPAFK 702
                    |::                    :|||  :  ::| ||: | | | : |
Db        450 ------AVE-------------------TVTEDAHTSLRLHRRGYTSAY-MRI-PQAA 480

Qy        703 GSAPINLSDRLHQVLRWALGSVEIFFSRHCPLWYGFGGGRLKWLQRLAYINTIVYPFTSL 762
              | |  :||  : | :||| | |:||   : ||   || | ||  ||| |:| : :  : :
Db        481 GLATESLSAHIGQRIRWARGMVQIFRLDN-PL---FGKG-LKLAQRLCYLNAMFHFLSGI 535

Qy        763 PLIAYCTIPAVCLLTGKFIIPTLSNLASMLFLGLFISIIVTAVLELRWSGVSIEDLWRNE 822
              | : : | |   ||  :||      | | : ||  :: ::    :  |:  :|:
Db        536 PRLIFLTAPLAFLLLHAYII-----YAPALMIALF---VIPHMVHASLTNSKIQGKYRH- 586

Qy        823 QFWVIGGVSAHLFAVFQGFLKMLAGIDTNFTVTAKAAEDAEFGELYMVKWTTLLIPPTTL 882
              ||                                        |:|       : ||| :
Db        587 SFW-------------------------------------SEIYETVLAWYIAPPTLV 607

Qy        883 LIINMSGCAGFSDALNKG------YEAW---GPLFGKVFFAFWVILHLYPFLKGL----M 929
              :||     |  :   ||      | | |     :  | |:|:|     |:
Db        608 ALINPH--KGKFNVTAKGGLVEEKYVDWVISRP------YIFLVLLNLLGVAAGVWRYYY 659

Qy        930 GRQNLTPTIVVLWSVLLASVFSLVWVKIN 958
              | :| | |::|          |||||  |
Db        660 GPENETLTVIV------- --SLVWVFYN 678


RESULT 3
BCA1_ACEXY
ID   BCA1_ACEXY     STANDARD;      PRT;   754 AA.
AC   P19449;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Cellulose synthase catalytic subunit [UDP-forming] (EC 2.4.1.12).
GN   BCSA.
OS   Acetobacter xylinus.
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;
OC   Acetobacteraceae; Gluconacetobacter.
OX   NCBI_TaxID=28448;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 25-42 AND 196-206.
RC   STRAIN=1306-3;
RX   MEDLINE=91045951; PubMed=2146681;
RA   Wong H.C., Fear A.L., Calhoon R.D., Eichinger G.H., Mayer R.,
RA   Amikam D., Benziman M., Gelfand D.H., Meade J.H., Emerick A.W.,
RA   Bruner R., Ben-Bassat A., Tal R.;
RT   "Genetic organization of the cellulose synthase operon in Acetobacter
RT   xylinum.";
```

Proc. Natl. Acad. Sci. U.S.A. 87:8130-8134(1990).
[2]
ENZYME REGULATION.
STRAIN=1306-3;
MEDLINE=21194671; PubMed=11297407;
Chang A.L., Tuckerman J.R., Gonzalez G., Mayer R., Weinhouse H., Volman G., Amikam D., Benziman M., Gilles-Gonzalez M. A.;
"Phosphodiesterase A1, a regulator of cellulose synthesis in Acetobacter xylinum, is a heme-based sensor.";
Biochemistry 40:3420-3426(2001).
-!- FUNCTION: Catalytic subunit of cellulose synthase. It polymerizes uridine 5'-diphosphate glucose to cellulose in a processive way. The thick cellulosic mats generated by this enzyme probably provide a specialized protective environment to the bacterium.
-!- CATALYTIC ACTIVITY: UDP-glucose + {(1,4)-beta-D-glucosyl}(N) = UDP + {(1,4)-beta-D-glucosyl}(N+1).
-!- COFACTOR: Magnesium.
-!- ENZYME REGULATION: Activated by bis-(3'-5') cyclic diguanylic acid (c-di-GMP).
-!- PATHWAY: Bacterial cellulose biosynthesis.
-!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane (Potential).
-!- INDUCTION: Cellulose is produced at a linear rate with respect to cell growth when O(2) is present.
-!- DOMAIN: There are two conserved domains in the globular part of the catalytic subunit: the N-terminal domain (domain A) contains the conserved DXD motif and is possibly involved in catalysis and substrate binding. The C-terminal domain (domain B) contains the QXXRW motif and is present only in processive glycosyl transferases. It could be involved in the processivity function of the enzyme, possibly required for holding the growing glycan chain in the active site.
-!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).

EMBL; M37202; AAA21884.1; -.
PIR; A43735; A43735.
InterPro; IPR003919; Cellsynth_A.
InterPro; IPR001173; Glyco_trans_2.
Pfam; PF00535; Glycos_transf_2; 1.
PRINTS; PR01439; CELLSNTHASEA.
Cellulose biosynthesis; Transferase; Glycosyltransferase; Transmembrane; Inner membrane.

DOMAIN	147	240	CATALYTIC SUBDOMAIN A.
DOMAIN	317	377	CATALYTIC SUBDOMAIN B.
TRANSMEM	26	46	POTENTIAL.
TRANSMEM	47	67	POTENTIAL.
TRANSMEM	108	128	POTENTIAL.
TRANSMEM	167	187	POTENTIAL.
TRANSMEM	407	427	POTENTIAL.
TRANSMEM	430	450	POTENTIAL.
TRANSMEM	468	488	POTENTIAL.
TRANSMEM	516	536	POTENTIAL.
TRANSMEM	549	569	POTENTIAL.
ACT_SITE	189	189	POTENTIAL.
ACT_SITE	333	333	POTENTIAL.
BINDING	236	236	SUBSTRATE (POTENTIAL).
BINDING	238	238	SUBSTRATE (POTENTIAL).

SEQUENCE 754 AA; 84442 MW; 8D5FC1FE62E2C068 CRC64;

Query Match 4.3%; Score 223; DB 1; Length 754;
Best Local Similarity 19.9%; Pred. No. 1.3e-08;
Matches 139; Conservative 67; Mismatches 210; Indels 284; Gaps 25;

```
Db        177 IIMRLVILGLFFHFRITNPVDSAFGLWL------TSVICEIWFAFSWVLDQFPKWNPVNR 230
```

```
Db         80 VLSALVVL RYLTW  I TLD  FDTWIQRGLSVTLLMAELYALYMLFLSYFQTIQPLHR 136

Qy        231 ETYIERLSAEYRRR  PSQLAGVDFFVSTVDPLKEPPLITANTVLSILAVDYPVDKVSCY 290

Db        137 --  --  APLPLP VVDDWPTVD FIPTYD  -EQLSIVRLTVLGALGIDWPPDKVNVY 185

Qy        291 VSDDGAAMLSFEAL TAEFARKWVPFCKKFSIEPRAPEFYFSQKIDYLKDKVQPSFVKE 350

Db        186 ILDDG  ---       --  -     --- - ------------      VRPEF -- 195

Qy        351 RRAMKREYEEFKVR LALVAKAQKTPEEGWTMQDGTPWPGNNTRDHPGHDSGLPWEIIGA 410

Db        196 ------ -EQFAKD AI            -  - ----------------------     205

Qy        411 RDIEGNELPRLVYV EPKPGYQHPKKAGAENALVRVSAVLTNAPYILNVDCDHYVNNSK 470

Db        206 ---  -   -  YI -N  -- ASPAKAGNLNDAIK  - RTSGDYILILDCDHI---  - 246

Qy        471 AVREAMTFLMDPQL  RDVCYVQFPQRFDGIDRHDRYANRNVVFFDVNMKGLDGIQGPVY 529

Db        247 PTRAFLQIAMGWT   --IAIMLTRENPWHPF          -- -- --- -- - --  273

Qy        530 VGTGCVFNRQALYG  GLAMPSLREKKDGSWFSSSDPSKKKPAQDFAEVYRDAKREDLN 589

Db        274 ------FQRNLAVS RTPPEGNL  --  - -  ----------------  -     293

Qy        590 AAIPNLTEICNYDP PRAMLIEQLSFEKTPSLRSVFIESTLMENGGVPESANSPPFIKEA 649

Db        294 --FYGVIQ DGND           FWDATFFDSCAI ----------------LRREA 318

Qy        650 IQVIGCGYEEKTEW FQIGWIYGSVTEDILSGFKMHCRGWRSIYCMPVRPAFKGSAPIN 709

Db        319 IESIG-  -          GFAVETVFEDAHTAIRMQRRGWSTAYLR--IPVASGLATPRI 363

Qy        710 SDRLHQVLRWALGSV EIFFSRHCPLWYGFGGGRLKWLQRLAYINTIVYPFTSLFLIAYCT 769

Db        364 TTHIGQRMRWARGM QIFRVDNPML-----GSGLKLGQRLCYLSAMTSFFFAIPRVIFLA 418

Qy        770 IPAVCLLTGKFIIP LSNLASMLFL  GLFISIVTAVLELRWSGVSIEDLWRNEQFW V 826

Db        419 SPLAFLFFGQNII-/ASFLAVLAYAIPHMFHFIATAAKVNKGW----------RYSFWSE 467

Qy        827 IGGVSAHLFAVFQG KYLAGIDTNFTVTAKAA EDAEF 864

Db        468 VYETTMALFLVRVT TLMPPSKGKFNVTEKG VIEEEEF 507
```

RESULT 4
BCSA_ECO57
ID BCSA_ECO57 STANDARD PRT; 872 AA.
AC Q8X5L7;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cellulose synthase catalytic subunit [UDP forming] (EC 2.4.1.12).
GN BCSA OR Z4948 OR ECS4412.
OS Escherichia coli O157:H7
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D., Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A., Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L., Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K., Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C., Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
RN [2]

```
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / RIMD 0509952;
RX   MEDLINE=21156231; PubMed=11258796;
RA   Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA   Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA   Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA   Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT   "Complete genome sequence of enterohemorrhagic Escherichia coli
RT   O157:H7 and genomic comparison with a laboratory strain K-12.";
RL   DNA Res. 8:11-22(2001).
CC   -!- FUNCTION: Catalytic subunit of cellulose synthase. It polymerizes
CC       uridine 5'-diphosphate glucose to cellulose, which is produced as
CC       an extracellular component for mechanical and chemical protection
CC       at the onset of the stationary phase, when the cells exhibit
CC       multicellular behavior (rdar morphotype). Co-expression of
CC       cellulose and thin aggregative fimbriae leads to a hydrophobic
CC       network with tightly packed cells embedded in a highly inert
CC       matrix (By similarity).
CC   -!- CATALYTIC ACTIVITY: UDP-glucose + {(1,4)-beta-D-glucosyl}(N) = UDP
CC       + {(1,4)-beta-D-glucosyl}(N+1).
CC   -!- COFACTOR: Magnesium (By similarity).
CC   -!- ENZYME REGULATION: Activated by bis-(3'-5') cyclic diguanylic acid
CC       (c-di-GMP) (By similarity).
CC   -!- PATHWAY: Bacterial cellulose biosynthesis.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC       (Potential).
CC   -!- DOMAIN: There are two conserved domains in the globular part of
CC       the protein: the N-terminal domain (domain A) contains
CC       the conserved DXD motif and is possibly involved in catalysis and
CC       substrate binding. The C-terminal domain (domain B) contains the
CC       QXXRW motif and is present only in processive glycosyl
CC       transferases. It could be involved in the processivity function of
CC       the enzyme, possibly required for holding the growing glycan chain
CC       in the active site.
CC   -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE005579; AAG58675.1; ALT_INIT.
DR   EMBL; AP002565; BAB37836.1; ALT_INIT.
DR   InterPro; IPR003919; Cellsynth_A.
DR   InterPro; IPR001173; Glyco_trans_2.
DR   Pfam; PF00535; Glycos_transf_2; 1.
DR   PRINTS; PR01439; CELLSNTHASEA.
KW   Cellulose biosynthesis; Transferase; Glycosyltransferase;
KW   Transmembrane; Inner membrane; Complete proteome.
FT   DOMAIN       271    364       CATALYTIC SUBDOMAIN A
FT   DOMAIN       441    501       CATALYTIC SUBDOMAIN B
FT   TRANSMEM      30     50       POTENTIAL.
FT   TRANSMEM     151    171       POTENTIAL.
FT   TRANSMEM     173    193       POTENTIAL.
FT   TRANSMEM     230    250       POTENTIAL.
FT   TRANSMEM     525    545       POTENTIAL.
FT   TRANSMEM     547    567       POTENTIAL.
FT   TRANSMEM     592    612       POTENTIAL.
FT   TRANSMEM     640    660       POTENTIAL.
FT   TRANSMEM     668    688       POTENTIAL.
FT   TRANSMEM     833    853       POTENTIAL.
FT   ACT_SITE     313    313       POTENTIAL.
FT   ACT_SITE     457    457       POTENTIAL.
FT   BINDING      360    360       SUBSTRATE (POTENTIAL)
FT   BINDING      362    362       SUBSTRATE (POTENTIAL)
SQ   SEQUENCE   872 AA;  99710 MW;  F3F1A24A2B713FBA CRC64;

  Query Match             4.2%;  Score 222;  DB 1;  Length 872;
  Best Local Similarity   17.8%;  Pred. No. 2e-08;
  Matches  145;  Conservative   87;  Mismatches  220;  Indels  362;  Gaps   28;

Qy        176 VIIMRLVILGLFFHFRIT------NPVDSAFGLWLTSVICEIWFAFSWVLDQFPKWNPVN 229
Db        202 LIVLSLTVSCRYIWWRYTSTLNWDDPVSLVCGLILLFAETYAWIVL--VLGYFQVVWPLN 259

Qy        230 RETYIERLSARYEREGEPSQLAGVDFFVSTVDPLKEPPLITANTVLSILAVDYPVDKVSC 289
Db        260 RQP--------VPLPKDMSLWPSVDIFVPTYN---EDLNVVKNTIYASLGIDWPKDKLNI 308

Qy        290 YVSDDGAAMLSFESLVETAEFARKWVPFCKKFSIEPRAPEFYFSQKIDYLKDKVQPSFVK 349
Db        309 WILDDGG---------------------------------------------------- 315

Qy        350 ERRAMKRDYEEYKVRVNALVAKAQKTPEEGWTMQDGTPWPGNNTRDHPGHDSGLPWEILG 409
Db        316 -----REEFRQFAQNVG------------------------------------------- 327

Qy        410 ARDIEGNELPRLVYVSREKRPGYQHHKKAGAENALVRVSAVLTNAPYILNVDCDHYVNNS 469
Db        328 ----------VKYIARTT---HEHAKAGNINNAL-----KYAKGEFVSIFDCDHVPTRS 368

Qy        470 KAVREAMCILMDPQVG-----RDVCYVQFPQRFDGIDKSDR-------YANRNVVFFDVN 517
Db        369 ---------FLQMTVGWFLKEKQLAMMQTPHHFFSPDPFERNLGRFRKTPNEGTLFYGLV 419

Qy        518 MKGLDGIQGPVYVGTGCVFNRQALYGYGPPSMPSLRKRKDSSSCFSCCCPSKKKPAQDPA 577
Db        420 QDGNDMWDATFFCGS------------------------------CAVIRRKP----- 442

Qy        578 EVYRDAKREDLNAAIFNLTEIDNYDEHERSMLISQLSFEKTFGLSSVFIESTLMENGGVP 637
Db        443 ---------------------------------------------------LDEIGGI- 449

Qy        638 ESANSPPFIKEAIQVIGCGYEEKTEWGKQIGWIYGSVTEDILSGFKMHCRGWRSIYCMPV 697
Db        450 -----------AVE--------------------TVTEDAHTSLRLHRRGYTSAY-MRI 476

Qy        698 RPAFKGSAPINLSDRLHQVLRWALGSVEIFFSRHCPLWYGFGGGRLKWLQRLAYINTIVY 757
Db        477 -PQAAGLATESLSAHIGQRIRWARGMVQIFRLDN-PL----TGKGLKFAQRLCYVNAMFH 530

Qy        758 PFTSLPLIAYCTIPAVCLLTGKFIIPTLSNLASMLFLGLFISIIVTAVLELRWSGVSIED 817
Db        531 FLSGIPRLIFLTAPLAFLLLHAYII-----YAPALMIALF---VLPHMIHASLTNSKIQG 582

Qy        818 LWRNEQFWVIGGVSAHLFAVFQGFLKMLAGIDTNFTVTAKAAEDAEFGELYMVKWTTLLI 877
Db        583 KYRH-SFW---------------------------------------SEIYETVLAWYIA 602

Qy        878 PPTTLLIINMSGCAGFSDALNKG---YEAWGPLFGKVFFAFWVILHLYPFLKGLMGRQNL 934
Db        603 PPTLVALINP----------HKGKFNVTAKGGLVEEE-YVDWVISRPYIFLVLL----NL 647

Qy        935 TPTIVVLW-------SVLLASVFSLVWVKINPFV 961
Db        648 VGVAVGIWRYFYGPPTEMLTVVVSMVWVFYNLIV 681


RESULT 5
BCSA_ECOLI
ID   BCSA_ECOLI     STANDARD;      PRT;   872 AA.
AC   P37653; P37654; P76712; P76713; Q8RSS7;
DT   01-OCT-1994 (Rel. 30, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Cellulose synthase catalytic subunit [UDP-forming] (EC 2.4.1.12).
GN   BCSA OR B3533.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
```

STRAIN=K12 / MG1655;
MEDLINE=94316500; PubMed=8041620;
Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;
"Analysis of the Escherichia coli genome. V. DNA sequence of the
region from 76.0 to 81.5 minutes.";
Nucleic Acids Res. 22:2576-2586(1994).
[2]
REVISIONS.
STRAIN=K12 / MG1655;
MEDLINE=97426617; PubMed=9278503;
Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
Mau B., Shao Y.;
"The complete genome sequence of Escherichia coli K-12.";
Science 277:1453-1474(1997).
[3]
CHARACTERIZATION.
STRAIN=ECOR10, ECOR12, and TOB1;
MEDLINE=21160181; PubMed=11260463;
Zogaj X., Nimtz M., Rohde M., Bokranz W., Roemling U.;
"The multicellular morphotypes of Salmonella typhimurium and
Escherichia coli produce cellulose as the second component of the
extracellular matrix.";
Mol. Microbiol. 39:1452-1463(2001).
-!- FUNCTION: Catalytic subunit of cellulose synthase. It polymerizes
uridine 5'-diphosphate glucose to cellulose, which is produced as
an extracellular component for mechanical and chemical protection
at the onset of the stationary phase, when the cells exhibit
multicellular behavior (rdar morphotype). Co-expression of
cellulose and thin aggregative fimbriae leads to a hydrophobic
network with tightly packed cells embedded in a highly inert
matrix.
-!- CATALYTIC ACTIVITY: UDP-glucose + {(1,4)-beta-D-glucosyl}(N) = UDP
+ {(1,4)-beta-D-glucosyl}(N+1).
-!- COFACTOR: Magnesium (By similarity).
-!- ENZYME REGULATION: Activated by bis-(3'-5') cyclic diguanylic acid
(c-di-GMP).
-!- PATHWAY: Bacterial cellulose biosynthesis.
-!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
(Potential).
-!- DOMAIN: There are two conserved domains in the globular part of
the protein: the N-terminal domain (domain A) contains the
conserved DXD motif and is possibly involved in catalysis and
substrate binding. The C-terminal domain (domain B) contains the
QXXRW motif and is present only in processive glycosyl
transferases. It could be involved in the processivity function of
the enzyme, possibly required for holding the growing glycan chain
in the active site.
-!- MISCELLANEOUS: The genes bscA, bcsB, bcsZ and bcsC are
constitutively transcribed but cellulose synthesis occurs only
when adrA, a putative transmembrane protein regulated by agfD, is
expressed. Cellulose production is abolished in E.coli K12.
-!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
-!- CAUTION: Ref.1 sequence differs from that shown due to a
frameshift in position 128.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).

DR EMBL; U00039; AAB18510.1; ALT_FRAME.
DR EMBL; U00039; AAB18511.1; ALT_FRAME.
DR EMBL; AE000430; AAC76558.1; ALT_INIT.
DR EcoGene; EG12260; bcsA.
DR InterPro; IPR003919; Cellsynth_A.
DR InterPro; IPR001173; Glyco_trans_2.
DR Pfam; PF00535; Glycos_transf_2; 1.
DR PRINTS; PR01439; CELLSNTHASEA.

KW Cellulose biosynthesis; Transferase; Glycosyltransferase;
KW Transmembrane; Inner membrane; Complete proteome.
FT DOMAIN 271 364 CATALYTIC SUBDOMAIN A.
FT DOMAIN 441 501 CATALYTIC SUBDOMAIN B.
FT TRANSMEM 30 56 POTENTIAL.
FT TRANSMEM 151 171 POTENTIAL.
FT TRANSMEM 173 193 POTENTIAL.
FT TRANSMEM 230 250 POTENTIAL
FT TRANSMEM 525 545 POTENTIAL.
FT TRANSMEM 547 567 POTENTIAL.
FT TRANSMEM 592 612 POTENTIAL.
FT TRANSMEM 640 660 POTENTIAL.
FT TRANSMEM 668 688 POTENTIAL.
FT TRANSMEM 833 853 POTENTIAL.
FT ACT_SITE 313 313 POTENTIAL.
FT ACT_SITE 457 457 POTENTIAL.
FT BINDING 360 360 SUBSTRATE (POTENTIAL).
FT BINDING 362 362 SUBSTRATE (POTENTIAL).
SQ SEQUENCE 872 AA; [illegible] MW; [illegible] CRC64;

Query Match [illegible]; Score 721; DB 1; Length 872;
Best Local Similarity [illegible]; Pred. No. 2.[illegible]e-08;
Matches 145; Conservative 86; Mismatches 226; Indels 352; Gaps 27;

```
Qy      176 VIIMRLVILGLFFHFRIT------NPVDSAFGLWLTSVICEIWFAFSWVLDQFPKWNPVN 229
            :|:: | :   :  :|.       :||    |; :        |    ||  |    ;:
Db      202 LIVLSLTVSCRYIWWRYTSTLNWDDPVSLVCGLILLFAETYAWIVL--VLGYFQVVWPLN 259

Qy      230 RETYIERLSARYERDGEPSQLAGVDFFVSTVDPLKEPPLITANTVLSILAVDYPVDKVSC 289
            |:              :  :    ,| ||   :   :   :  ||: : | :|:| :|:: 
Db      260 RQP--  -----VPLPKDMSLWPSVDIFVPTYN---EDLNVVKNTIYASLGIDWPKDKLNI 308

Qy      290 YVSDDGAAMLSFESLVETAEFARKWVPFCKKFSIEPRAPEFYFSQKIDYLKDKVQPSFVK 349
            :: :||
Db      309 WILDDGG----  ------------------    -- ---------------------  - 345

Qy      350 ERRAMKRDYEEYKVRVNALVAKAQKTPEEGWTMQDGTPWPGNNTRDHPGHDSGLPWEILG 409
                : :: ::   |
Db      316 -----REEFRQFAQNVG-------------------------------  --  ----    327

Qy      410 ARDIEGNELPRLVYVSREKRPGYQHHKKAGAENALVRVSAVLTNAPYILNVDCDHYVNNS 469
                    : |::'    ::| |     !!|           ::   |.!!    :
Db      328 ----------VKYIARTT---HEHAKAGNINNALL----KYAKGEFVSIFDCDHVPTRS 368

Qy      470 KAVREAMCILMDPQVGRDVCYVQFPQRFDGIDKSDR-------YANRNVVFFDVNMKGLD 522
                  | : |    :  :| |  :   |  :|         |   :|: :   ' '
Db      369 FLQMTMGWFLKEKQ---LAMMQTPHHFFSPDPFERNLGRFRKTPNEGTLFYGLVQDGND 424

Qy      523 GIQGPVYVGTGCVFNRQALYGYGPPSMPSLRKRKDSSSCFSCCCPSKKKPAQDPAEVYRD 582
                : |:                                    :  ::.|
Db      425 MWDATFFCGS-------------------     ---------CAVIRRKP---------- 442

Qy      583 AKREDLNAAIFNMEDNYDEHERSMLISQLSFEKTFGLSSVFIESTLMENGGVPESANS 642
                                                          | | ||:
Db      443 ------------- ---------    - -   ------------LDEIGGI------ 449

Qy      643 PPFIKEAIQVIGCGYEEKTEWGKQIGWIYGSVTEDILSGFKMHCRGWRSIYCMPVRPAFK 702
                |::                    :.|||   :  ::| ||:  | | : :
Db      450 ------AVE---------------- -- TVTEDAHTSLRLHRRGYTSAY-MRI-PQAA 480

Qy      703 GSAPINLSDRLHQVLRWALGSVEIFFSRHCPLWYGFGGGRLKWLQRLAYINTIVYPFTSL 762
            | |  :||  : | :|||  | |:||    : ||     |  ||: |||  |:| : :  : :
Db      481 GLATESLSAHIGQRIRWARGMVQIFRLDN PL -  TGKGLKFAQRLCYVNAMFHFLSGI 535

Qy      763 PLIAYCTIPAVCLLTGKFIIPTLSNLASMLFLGLFISIIVTAVLELRWSGVSIEDLWRNE 822
            | : : | |   ||   :||       |  | : ||   ::  ::     :   |:  :|:
Db      536 PRLIFLTAPLAFLLLHAYII-----YAPALMIALF---VLPHMIHASLTNSKIQGKYRH- 586

Qy      823 QFWVIGGVSAHLFAVFQGFLKMLAGIDTNFTVTAKAAEDAEFGELYMVKWTTLLIPPTTL 882
            ||                                        |:|      : |:| :
Db      587 SFW-----------------------------------------SEIYETVLAWYIAPPTLV 607

Qy      883 LIINMSGCAGFSDALNKG---YEAWGPLFGKVFFAFWVILHLYPFLKGLMGRQNLTPTIV 939
```

```
            :||          :||     | | |  :  :  |||   | ||  |    ||    |
Db        608 ALINP----------HKGKFNVTAKGGLVEEE-YVDWVISRPYIFLVLL----NLVGVAV 652

Qy        940 VLW-------SVLLASVFSLVWVKINPFV 961
              :|       : :|  | |:|||  |  |
Db        653 GIWRYFYGPPTEMLTVVVSMVWVFYNLIV 681


RESULT 6
BCA2_ACEXY
ID   BCA2_ACEXY     STANDARD;      PRT;   756 AA.
AC   O82859;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Cellulose synthase catalytic subunit [UDP-forming] (EC 2.4.1.12).
GN   BCSA.
OS   Acetobacter xylinus.
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;
OC   Acetobacteraceae; Gluconacetobacter.
OX   NCBI_TaxID=28448;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BPR 2001;
RX   MEDLINE=98296257; PubMed=9630539;
RA   Nakai T., Moriya A., Tonouchi N., Tsuchida T., Yoshinaga F.,
RA   Horinouchi S., Sone Y., Mori H., Sakai F., Hayashi T.;
RT   "Control of expression by the cellulose synthase (bcsA) promoter
RT   region from Acetobacter xylinum BPR 2001.";
RL   Gene 213:93-100(1998).
CC   -!- FUNCTION: Catalytic subunit of cellulose synthase. It polymerizes
CC       uridine 5'-diphosphate glucose to cellulose. The thick cellulosic
CC       mats generated by this enzyme probably provide a specialized
CC       protective environment to the bacterium (By similarity).
CC    !  CATALYTIC ACTIVITY: UDP-glucose + {(1,4)-beta-D-glucosyl}(N) = UDP
CC       + {(1,4)-beta-D-glucosyl}(N+1).
CC   -!- COFACTOR: Magnesium (By similarity).
CC   -!- ENZYME REGULATION: Activated by bis-(3'-5') cyclic diguanylic acid
CC       (c-di-GMP).
CC   -!- PATHWAY: Bacterial cellulose biosynthesis.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC       (Potential).
CC   -!- DOMAIN: There are two conserved domains in the globular part of
CC       the catalytic subunit: the N-terminal domain (domain A) contains
CC       the conserved DXD motif and is possibly involved in catalysis and
CC       substrate binding. The C terminal domain (domain B) contains the
CC       QXXRW motif and is present only in processive glycosyl
CC       transferases. It could be involved in the processivity function of
CC       the enzyme, possibly required for holding the growing glycan chain
CC       in the active site.
CC   -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AB010645; BAA31463.1; -.
DR   InterPro; IPR003919; Cellsynth_A.
DR   InterPro; IPR001173; Glyco_trans_2.
DR   Pfam; PF00535; Glycos_transf_2; 1.
DR   PRINTS; PR01439; CELLSNTHASEA.
KW   Cellulose biosynthesis; Transferase; Glycosyltransferase;
KW   Transmembrane; Inner membrane.
FT   DOMAIN      147    242       CATALYTIC SUBDOMAIN A.
FT   DOMAIN      319    379       CATALYTIC SUBDOMAIN B.
FT   TRANSMEM     27     47       POTENTIAL.
FT   TRANSMEM     49     69       POTENTIAL.
FT   TRANSMEM    106    126       POTENTIAL.
FT   TRANSMEM    167    187       POTENTIAL.
FT   TRANSMEM    409    429       POTENTIAL.
FT   TRANSMEM    432    452       POTENTIAL.
FT   TRANSMEM    470    490       POTENTIAL.
FT   TRANSMEM    517    537       POTENTIAL.
FT   TRANSMEM    551    571       POTENTIAL.
FT   ACT_SITE    189    189       POTENTIAL.
FT   ACT_SITE    335    335       POTENTIAL.
FT   BINDING     238    238       SUBSTRATE (POTENTIAL).
FT   BINDING     240    240       SUBSTRATE (POTENTIAL).
SQ   SEQUENCE   756 AA;  84561 MW;  6954F39A25E73B0A CRC64;

  Query Match             4.1%;  Score 217;  DB 1;  Length 756;
  Best Local Similarity   19.5%;  Pred. No. 3.8e-08;
  Matches  136;  Conservative   68;  Mismatches  215;  Indels  280;  Gaps   24;

Qy        177 IIMRLVILGLFFHFRITNPVDSAFGLWLTSVI------CEIWFAFSWVLDQFPKWNPVNR 230
              ::  || |  :  :|:|  .|  |  |:  ::       |::  :   |  |    |::|
Db         80 VLSALVSL-RYLTWRLTETLD--FNTWIQGILGVILLMAELYALYMLFLSYFQTIQPLHR 136

Qy        231 ETYIERLSARYEREGEPSQLAGVDFFVSTVDPLKEPPLITANTVLSILAVDYPVDKVSCY 290
                      |              || |: | |   |   |   |||  | :|:| |||: |
Db        137 --------APLPLPDNVDDWPTVDIFIPTYD---EQLSIVRLTVLGALGIDWPPDKVNVY 185

Qy        291 VSDDGAAMLSFESLVETAEFARKWVPFCKKFSIEPRAPEFYFSQKIDYLKDKVQPSFVKE 350
              : |||                                                |:| |
Db        186 ILDDG-------------------------------------------------VRPEF--- 195

Qy        351 RRAMKRDYEEYKVRVNALVAKAQKTPEEGWTMQDGTPWPGNNTRDHPGHDSGLPWEILGA 410
                    |::     ||                                        :|
Db        196 --------EQFAKDCGAL-----------------------------------YIGR 209

Qy        411 RDIEGNELPRLVYVSREKRPGYQHHKKAGAENALVRVSAVLTNAPYILNVDCDHYVNNSK 470
               |::                    | |||  |  ::     |:  ||| :||||
Db        210 VDVD-----------------SAHAKAGNLNHAIK----RTSGDYILILDCDHI----- 242

Qy        471 AVREAMCILMDPQVG-RDVCYVQFPQRFDGIDKSDRYANRNVVFFDVNMKGLDGIQGPVY 529
                | : | |   | | :  :| | |  |   |
Db        243 PTRAFLQIAMGWMVADRKIALMQTPHHFYSPDP-------------------------- 275

Qy        530 VGTGCVFNRQALYGYGPPSMPSLRKRKDSSSCFSCCCPSKKKPAQDPAEVYRDAKREDLN 589
                  | |    ||  |   :|
Db        276 ------FQRNLAVGYRTPPEGNL----------------------------------- 292

Qy        590 AAIFNLTEIDNYDEHERSMLISQLSFEKTFGLSSVFIESTLMENGGVPESANSPPFIKEA 649
                 : : : |  |             :: ||   |  |                    :||
Db        293 --FYGVIQ-DGND-----------FWDATFFCGSCAI----------------LRREA 320

Qy        650 IQVIGCGYEEKTEWGKQIGWIYGSVTEDILSGFKMHCRGWRSIYCMPVRPAFKGSAPINL 709
              |: ||              |:   :|||  :  :|  ||| : |     |   | |   !
Db        321 IESIG------------GFAVETVTEDAHTALRMQRRGWSTAYLR--IPVASGLATERL 365

Qy        710 SDRLHQVLRWALGSVEIFFSRHCPLWYGFGGGRLKWLQRLAYINTIVYPFTSLPLIAYCT 769
              :  : | :||| | ::||   :  |     |  ||  ||| |:: :   | ::| : :
Db        366 TTHIGQRMRWARGMIQIFRVDNPML-----GRGLKLGQRLCYLSAMTSFFFAIPRVIFLA 420

Qy        770 IPAVCLLTGKFIIPTLS-NLASMLFLGLFISIIVTAVLELRWSGVSIEDLWRNEQFW-VI 827
               |   | |: ||       :|:     :| ||  | :   |                || :
Db        421 SPLAFLFAGQNIIAAAPLAVAAYALPHMFHSIATAAKVNKGW----------RYSFWSEV 470

Qy        828 GGVSAHLFAVFQGFLKMLAGIDTNFTVTAKAA--EDAEF 864
               :  || |    : :|      | || |    |: ||
Db        471 YETTMALFLVRVTIVTLLFPSKGKFNVTEKGGVLEEEEF 509


RESULT 7
BCSA_PSEFL
ID   BCSA_PSEFL     STANDARD;      PRT;   739 AA.
AC   P58931; Q8RSZ1;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Cellulose synthase catalytic subunit [UDP-forming] (EC 2.4.1.12).
GN   BCSA OR WSSB.
```

```
S   Pseudomonas fluorescens.
C   Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
C   Pseudomonadaceae; Pseudomonas.
X   NCBI_TaxID=294;
N   [1]
P   SEQUENCE FROM N.A.
C   STRAIN=SBW25;
X   MEDLINE=22013850; PubMed=12019221;
A   Spiers A.J., Kahn S.G., Bohannon J., Travisano M., Rainey P.B.;
T   "Adaptive divergence in experimental populations of Pseudomonas
T   fluorescens. I. Genetic and phenotypic bases of wrinkly spreader
T   fitness.";
L   Genetics 161:33-46(2002).
C   -!- FUNCTION: Catalytic subunit of cellulose synthase. It polymerizes
C       uridine 5'-diphosphate glucose to cellulose, which is produced as
C       an extracellular component responsible for the structural
C       integrity and rigidity of self-supporting mats characteristic of
C       the "wrinkly spreader" phenotype.
C   -!- CATALYTIC ACTIVITY: UDP-glucose + {(1,4)-beta-D-glucosyl}(N)   UDP
C       + {(1,4)-beta-D-glucosyl}(N+1).
C   -!- COFACTOR: Magnesium (By similarity).
C   -!- ENZYME REGULATION: Activated by bis-(3'-5') cyclic dimanylic acid
C       (c-di-GMP) (By similarity).
C   -!- PATHWAY: Bacterial cellulose biosynthesis.
C   -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
C       (Potential).
C   -!- DOMAIN: There are two conserved domains in the globular part of
C       the protein: the N-terminal domain (domain A) contains
C       the conserved DXD motif and is possibly involved in catalysis and
C       substrate binding. The C-terminal domain (domain B) contains the
C       QXXRW motif and is present only in processive glycosyl
C       transferases. It could be involved in the processivity function of
C       the enzyme, possibly required for holding the growing glycan chain
C       in the active site.
C   -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
C   ----------------------------------------------------------------------
C   This SWISS-PROT entry is copyright. It is produced through a collaboration
C   between the Swiss Institute of Bioinformatics and the EMBL outstation -
C   the European Bioinformatics Institute. There are no restrictions on its
C   use by non-profit institutions as long as its content is in no way
C   modified and this statement is not removed. Usage by and for commercial
C   entities requires a license agreement (See http://www.isb-sib.ch/announce/
C   or send an email to license@isb-sib.ch).
C   ----------------------------------------------------------------------
R   EMBL; AY074776; AAL71842.1; -.
R   InterPro; IPR001173; Glyco_trans_2.
R   Pfam; PF00535; Glycos_transf_2; 1.
W   Cellulose biosynthesis; Transferase; Glycosyltransferase;
W   Transmembrane; Inner membrane.
T   DOMAIN      157     250       CATALYTIC SUBDOMAIN A.
T   DOMAIN      327     387       CATALYTIC SUBDOMAIN B.
T   TRANSMEM     36      55       POTENTIAL.
T   TRANSMEM     59      76       POTENTIAL.
T   TRANSMEM     83     101       POTENTIAL.
T   TRANSMEM    116     138       POTENTIAL.
T   TRANSMEM    417     436       POTENTIAL.
T   TRANSMEM    440     462       POTENTIAL.
T   TRANSMEM    524     546       POTENTIAL.
T   TRANSMEM    551     573       POTENTIAL.
T   ACT_SITE    199     199       POTENTIAL.
T   ACT_SITE    343     343       POTENTIAL.
T   BINDING     246     246       SUBSTRATE (POTENTIAL).
T   BINDING     248     248       SUBSTRATE (POTENTIAL).
Q   SEQUENCE   739 AA;  82165 MW;  2B962EA3854B23BB CRC64;

 Query Match             4.1%;  Score 214;  DB 1;  Length 739;
 Best Local Similarity   19.6%;  Pred. No. 6.1e-08;
 Matches  166;  Conservative   89;  Mismatches  232;  Indels  360;  Gaps   37;

Qy       173 YRAVIIMR----------LVILGL-----FFHFRITNP------VDSAFGLWLTS----V 207
             : ||:::|          ||:| |     :  :|:|:       ||  ||  | :
Db        70 FLAVLVLRKIPGRLAILALVVLSLVASLRYMFWRLTSTLGFETWVDMFFGYGLVAAEFYA 129

Qy       208 ICRIWFAF---SWV QQFPKWNPVNRETYIERLSARYEREGEPSQLAGVDFFVSTVDPLK 264
             :  :   :   :   : | |                    : :| :   || | : :
Db       130 LIVLIFGYVQTAWFLRRIPVW---  ------      --LKTEPEEWPTVDVFIPTYN -  169

Qy       265 EPPLITANTVLSILAVDYPVDKVSCYVSDDGAAMLSFESLVETAEFARKWVPFCKKFSIE 324
             |   |   |: :  :.:: ||:  :| ||
Db       170 EALSIVKHTIFAAQAMDWPKDKIRVHVLDDG ---------------- -----  --- 200

Qy       325 PRAPEFYFSQKIDY KDKVQPSFVKERRAMKREYEEY--KVRVNALVAKAQKTPEEGWTM 382
                                          : |: |:  || ||
Db       201 ---  - --  -          --   -   --RRLDFREFCRKVGVN---------------- 215

Qy       383 QDGTPWPGNNTRDH GHPSGLLFWRILGARDIEGNELPRLVYVSREKRPGYQHHKKAGAEN 442
                                                    |: |:       : |||  '
Db       216 ----  - ---              -- -------------YIRRDN   -NFHAKAGNLN 231

Qy       443 ALVRVSAVLTNAPY LNVDCDHYVNNSKAVREAMCFLMDPQVGRDVCYVQFPQRFDGIDK 502
               ::|    |:  :    | |    :          | |:    :  :|    |   |
Db       232 EALE7    FFRPF S  A.RLFT  FIQVFLWFLKDPK -- LAMLQTPHFFFSPDP 283

Qy       503 SDRYANRNVVFFI  MALDPLFGFSTVGYFWNFQALYGYGPPSMPSLRKPKDSSSCF 562
                            |  |  :          :  |
Db       284 FE   --------   KNLDTFRAVPNEG--    --ELFYGL----  -  -------- 304

Qy       563 SCCCPSKKKPAQDFAEVYRDAKRCLNAAIFNLTEIDNYDEHERSMLISQLSFEKTFGLS 622
                       |:  :::         ||  :                          |
Db       305 -----  --  VQDGNDLW--------NATFF    ---------------------CGSC 321

Qy       623 SVFIESTLMENGGVIESANSPPFIKEAIQVIGCGYEEKTEWGKQIGWIYGSVTEDILSGF 682
             :|      |:| |||           |::                          :||||  :
Db       322 AVIRREPLEIGGV -- ------- AVE       --------------TVTEDAHTAL 348

Qy       683 KMHCRGWRSIYCMPVRPAFKGSAPINLSDRLHQVLRWALGSVEIFFSRHCPLWYGFGGGR 742
             |::  ::  :  |          :  |  :|:  :::  :|||  |  :||  :  |:    :
Db       349 KLNRLGYNTAYL- AIPCAAGLATESLSRHINQRIRWARGMAQIFRTDN-PLL----GKC 401

Qy       743 LKWLQRLAYINTIVYPFTSLPLIAYCTIPAVCLLTGKFIIPTLSNLASMLFLGLFI- SI 800
             |||  |:  | |  :  :  ||  :  :  '    :  |  |       |  |  :  ::  :
Db       402 LKWGQRICYANAMCHFFYGLPRLVFLTAPLAYLIFGAEIFH-----ASALMIVAYVLPHL 456

Qy       801 IVTAVLELRWSGVSIEDLWRNEQF   --WVIGGVSAHLFAVFQGFLKMLAGIDTNFTVT 858
             : :::   |  |      | ||  :   | |      |  :   :   ||    | |:
Db       457 VHSSLTNSRIQGRFRHSFW-NEVYETVLAWYI------LPPVLVALVNPKAG---GFNVT 506

Qy       856 AKAA-EDAEFGELYMVKWTTLLIPPTTLLINMSGCA-GFSDALNKGYEAWGPLFGKVFF 913
               |   | :: :    |  :  |  || :: |   |      :      ||     |
Db       507 DKGGIIDKQFFD   --W-KLARPYLVLLAVNLIGLGFGIHQLI--  ---WGDASTAVTV 554

Qy       914 AFWVILHLYPFLKGLMGRQNLTPTI  VVLWSVLLASVFSLVWVKINPFVNKVDNTLVAE 971
             :                  :||  |:  ::: ,  :|      |:  |  |:       |:
Db       555 AI------ ----- ----NLTWTLYNLIITSAAVAVASEARQVRSEPRVS-  ----AK 590

Qy       972 TCISIDC 978
               :||
Db       591 LPVSIIC 597


RESULT 8
BCA4_ACEXY
ID   BCA4_ACEXY     STANDARD;      PRT;  1518 AA.
AC   Q9RBJ2;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Putative cellulose synthase 2 [Includes: Cellulose synthase
DE   catalytic subunit [UDP-forming] (EC 2.4.1.12); Cyclic di-GMP binding
DE   domain (Cellulose synthase 2 regulatory subunit)].
GN   BCSABII-A.
OS   Acetobacter xylinus.
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;
OC   Acetobacteraceae; Gluconacetobacter.
OX   NCBI_TaxID=28448;
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RC   STRAIN=JCM 7664 / IFO 13693;
RX   MEDLINE=99310341; PubMed=10382968;
RA   Umeda Y., Hirano A., Ishibashi M., Akiyama H., Onizuka T., Ikeuchi M.,
RA   Inoue Y.;
RT   "Cloning of cellulose synthase genes from Acetobacter xylinum JCM
RT   7664: implication of a novel set of cellulose synthase genes.";
RL   DNA Res. 6:109-115(1999).
CC   -!- CATALYTIC ACTIVITY: UDP-glucose + {(1,4)-beta-D-glucosyl}(N) = UDP
CC       + {(1,4)-beta-D-glucosyl}(N+1).
CC   -!- PATHWAY: Bacterial cellulose biosynthesis.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC       (Potential).
CC   -!- DOMAIN: There are two conserved domains in the globular part of
CC       the catalytic subunit: the N terminal domain (domain A) contains
CC       the conserved DXD motif and is possibly involved in catalysis and
CC       substrate binding. The C-terminal domain (domain B) contains the
CC       QXXRW motif and is present only in processive glycosyl
CC       transferases. It could be involved in the processivity function of
CC       the enzyme, possibly required for holding the growing glycan chain
CC       in the active site.
CC   -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE
CC       GLYCOSYLTRANSFERASE FAMILY 2.
CC   -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE ACSB/BCSB
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB015803; BAA77593.1; -.
DR   InterPro; IPR003919; Cellsynth_A.
DR   InterPro; IPR003920; Cellsynth_B.
DR   InterPro; IPR001173; Glyco_trans_2.
DR   Pfam; PF03170; BcsB; 1.
DR   Pfam; PF00535; Glycos_transf_2; 1.
DR   PRINTS; PR01439; CELLSNTHASEA.
DR   PRINTS; PR01440; CELLSNTHASEB.
KW   Cellulose biosynthesis; Transferase; Glycosyltransferase;
KW   Transmembrane; Inner membrane.
FT   DOMAIN          1    731       CATALYTIC.
FT   DOMAIN        732   1518       CYCLIC DI-GMP BINDING (BY SIMILARITY).
FT   DOMAIN        144    237       CATALYTIC SUBDOMAIN A.
FT   DOMAIN        314    374       CATALYTIC SUBDOMAIN B.
FT   TRANSMEM       24     44       POTENTIAL.
FT   TRANSMEM       71     91       POTENTIAL.
FT   TRANSMEM      105    125       POTENTIAL.
FT   TRANSMEM      404    424       POTENTIAL.
FT   TRANSMEM      427    447       POTENTIAL.
FT   TRANSMEM      465    485       POTENTIAL.
FT   TRANSMEM      514    534       POTENTIAL.
FT   TRANSMEM      543    563       POTENTIAL.
FT   TRANSMEM     1481   1501       POTENTIAL.
FT   ACT_SITE      186    186       POTENTIAL.
FT   ACT_SITE      330    330       POTENTIAL.
FT   BINDING       233    233       SUBSTRATE (POTENTIAL).
FT   BINDING       235    235       SUBSTRATE (POTENTIAL).
SQ   SEQUENCE   1518 AA;  166404 MW;  0EC99B35B6DE4543 CRC64;

  Query Match             4.0%;  Score 211.5;  DB 1;  Length 1518;
  Best Local Similarity  16.3%;  Pred. No. 2.7e-07;
  Matches  131;  Conservative   86;  Mismatches  216;  Indels  371;  Gaps   20;

Qy        176 VIIMRLVILGLFFHFRITNPVDSAFGLWLTS--  -----   ---VICEIWFAFSWVLD 220
              :::: |::   :  :|:|  |   '  || :                :|  :|  :|
Db         75 LMMLSLLVSLRYIVWRLTATVQ- FSNWLQTALAVLLLAEIYALMTLCLSYFQMAW--- 129

Qy        221 QFPKWNPVNRETYIERLSARYERFGEPSQLAGVDFFVSTVDPLKEPPLITANTVLSILAV 280
                  |: |         :    : : :   || || : :   |   :  :|||  | :

Db        130 ------PLRRR--------EHPLPEDMAQWPSVDVFVPSYN---EELSLVRSTVLGALDL 172

Qy        281 DYPVDKVSCYVSDDGAAMLSFESLVETAEFARKWVPFCKKFSIEPRAPEFYFSQKIDYLK 340
              |:| |::: |: |||
Db        173 DWPADRLNVYILDDG-- ---------------------------------------- 187

Qy        341 DKVQPSFVKERRAMKRDYEEYKVRVNALVAKAQKTPEEGWTMQDGTPWPGNNTRDHPGHD 400
                        :: : :: |   |
Db        188 -------------RRKAFHDFAVEAGA---------------------------------- 201

Qy        401 SGLPWEILGARDIEGNELPRLVYVSREKRPGYQHHKKAGAENALVRVSAVLTNAPYILNV 460
                                  |: | :    :| |||  |  : |    |::|: :
Db        202 ----------------- --GYIIRAE----NNHAKAGNLNHALAV----TDSPFAVIF 232

Qy        461 DCDHYVNNSKAVREAMCILMDPQVGRDVCYVQFPQRFDGIDKSDRYANRNV-------VF 513
              ||||         |    :: ||    ::  :| |  |   |   |     :       :|
Db        233 DCDHVPTRGFLRRTIGWMMADP----NLALLQTPHHFYAPDPFQRNLAGGMHVPPEGNMF 288

Qy        514 FDVNMKGLDGIQGPVYVGTGCVFNRQALYGYGPPSMPSLRKRKDSSSCFSCCCPSKKKPA 573
              : :   | |      : |:  :  |:|: | |
Db        289 YGLVQDGNDFWDATFFCGSCAIIRREAVMGIG---------------------------- 320

Qy        574 QDPAEVYRDAKREDLNAAIFNLTEIDNYDEHERSMLISQLSFEKTFGLSSVFIESTLMEN 633

Db        321 ------------------------------------------------------------ 320

Qy        634 GGVPESANSPPFIKEAIQVIGCGYEEKTEWGKQIGWIYGSVTEDILSGFKMHCRGWRSIY 693
                                                |:   :|||  :  ||  ||| : |
Db        321 ----------------- ----------------GFATETVTEDAHTALKMQRRGWGTAY 346

Qy        694 CMPVRPAFKGSAPINLSDRLHQVLRWALGSVEIFFSRHCPLWYGFGGGRLKWLQRLAYIN 753
                  |  | |   |   : | :||| | ::|    :  |     |  |:| ||| |::
Db        347 LR--EPLAAGLATERLILHIGQRVRWARGMIQIMRLDNPML-----GAGLRWEQRLCYLS 399

Qy        754 TIVYPFTSLPLIAYCTIPAVCLLTGKFIIPTLSNLASMLFL--GLFISIIVTAVLELRWS 811
                : :   ::| : :   |   |  |: ||   | ||  ::    :| |:|  : :| ||
Db        400 AMSHFLFAIPRLTFLVSPLAFLFLGQNII-AASPLAISVYALPHIFHSVITLSRIEGRW- 457

Qy        812 GVSIEDLWRNEQFW-VIGGVSAHLFAVFQGFLKMLAGIDTNFTVTAKAAEDAEFGELYMV 870
                        ||  |   |  || |  : :|       | || |
Db        458 ---------RYSFWSEIYETSLALFLVRITIVTLLQPHKGKFNVTDKGG----------- 497

Qy        871 KWTTLLIPPTTLLIINMSGCAGFSDALNKGYEAWGPLFGKVFFAFWVILHLYPFLKGLMG 930
                                      | :||  |  ::  |  |
Db        498 ---------- ----------------LLARGYFDWDAVYPNVILA---------------- 516

Qy        931 RQNLTPTIVVLWSVLLASVFSLVW 954
                    '| : ||  |  :||
Db        517 ------  GVLCAALLRGVPGIVW 532


RESULT 9
BCA5_ACEXY
ID   BCA5_ACEXY     STANDARD;      PRT;  1518 AA.
AC   Q9WX75;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Putative cellulose synthase 3 [Includes: Cellulose synthase
DE   catalytic subunit [UDP-forming] (EC 2.4.1.12); Cyclic di-GMP binding
DE   domain (Cellulose synthase 3 regulatory subunit)].
GN   BCSABII-B.
OS   Acetobacter xylinus.
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;
OC   Acetobacteraceae; Gluconacetobacter.
OX   NCBI_TaxID=28448;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JCM 7664 / IFO 13693;
RX   MEDLINE=99310341; PubMed=10382968;
RA   Umeda Y., Hirano A., Ishibashi M., Akiyama H., Onizuka T., Ikeuchi M.,
RA   Inoue Y.;
RT   "Cloning of cellulose synthase genes from Acetobacter xylinum JCM
```

```
T   7664: implication of a novel set of cellulose synthase genes.";
L   DNA Res. 6:109-115(1999).
C   -!- CATALYTIC ACTIVITY: UDP-glucose + {(1,4)-beta-D-glucosyl}(N) = UDP
C       + {(1,4)-beta-D-glucosyl}(N+1).
C   -!- PATHWAY: Bacterial cellulose biosynthesis.
C   -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
C       (Potential).
C   -!- DOMAIN: There are two conserved domains in the globular part of
C       the catalytic subunit: the N-terminal domain (domain A) contains
C       the conserved DXD motif and is possibly involved in catalysis and
C       substrate binding. The C-terminal domain (domain B) contains the
C       QXXRW motif and is present only in processive glycosyl
C       transferases. It could be involved in the processivity function of
C       the enzyme, possibly required for holding the growing glycan chain
C       in the active site.
C   -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE
C       GLYCOSYLTRANSFERASE FAMILY 2.
C   -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE ACSB/BCSB
C       FAMILY.
C   ---------------------------------------- ----------------------------
C   This SWISS-PROT entry is copyright. It is produced through a collaboration
C   between the Swiss Institute of Bioinformatics and the EMBL outstation -
C   the European Bioinformatics Institute. There are no restrictions on its
C   use by non-profit institutions as long as its content is in no way
C   modified and this statement is not removed. Usage by and for commercial
C   entities requires a license agreement (See http://www.isb-sib.ch/announce/
C   or send an email to license@isb-sib.ch).
C   ---------------------------------------- ----------------------------
R   EMBL; AB015804; BAA77600.1; -.
R   InterPro; IPR003919; Cellsynth_A.
R   InterPro; IPR003920; Cellsynth_B.
R   InterPro; IPR001173; Glyco_trans_2.
R   Pfam; PF03170; BcsB; 1.
R   Pfam; PF00535; Glycos_transf_2; 1.
R   PRINTS; PR01439; CELLSNTHASEA.
R   PRINTS; PR01440; CELLSNTHASEB.
W   Cellulose biosynthesis; Transferase; Glycosyltransferase;
W   Transmembrane; Inner membrane.
T   DOMAIN        1    731       CATALYTIC.
T   DOMAIN      732   1518       CYCLIC DI-GMP BINDING (BY SIMILARITY).
T   DOMAIN      144    237       CATALYTIC SUBDOMAIN A.
T   DOMAIN      314    374       CATALYTIC SUBDOMAIN B.
T   TRANSMEM     24     44       POTENTIAL.
T   TRANSMEM     71     91       POTENTIAL.
T   TRANSMEM    105    125       POTENTIAL.
T   TRANSMEM    404    424       POTENTIAL.
T   TRANSMEM    428    448       POTENTIAL.
T   TRANSMEM    465    485       POTENTIAL.
T   TRANSMEM    514    534       POTENTIAL.
T   TRANSMEM    543    563       POTENTIAL.
T   TRANSMEM   1481   1501       POTENTIAL.
T   ACT_SITE    186    186       POTENTIAL.
T   ACT_SITE    330    330       POTENTIAL.
T   SITE        233    233       IMPORTANT FOR SUBSTRATE BINDING
T                                (POTENTIAL).
T   SITE        235    235       IMPORTANT FOR SUBSTRATE BINDING
T                                (POTENTIAL).
Q   SEQUENCE   1518 AA;  166464 MW;  7D76345031B3DAB6 CRC64;

 Query Match             4.0%;  Score 211.5;  DB 1;  Length 1518;
 Best Local Similarity   16.3%;  Pred. No. 2.7e-07;
 Matches  131;  Conservative   86;  Mismatches  216;  Indels  371;  Gaps   20;

y       176 VIIMRLVILGLFFHFRITNPVDSAFGLWLTS--------------VICEIWFAFSWVLD 220
            :::: |::   :  :|:|  |   |  || :              :|  :|  :|
b        75 LMMLSLLVSLRYIVWRLTATVQ--FSNWLQTALAVLLLLAEAYALMTLCLSYFQMAW--- 129

y       221 QFPKWNPVNRETYIERLSARYEREGEPSQLAGVDFFVSTVDPLKEPPLITANTVLSILAV 280
                  |: |          :    : :|   || || : :   |   :  :|||  | :
b       130 ------PLRRR--------EHPLPEDMAQWPSVDVFVPSYN---EELSLVRSTVLGALDL 172

y       281 DYPVDKVSCYVSDDGAAMLSFESLVETAEFARKWVPFCKKFSIEPRAPEFYFSQKIDYLK 340
            |:| |::: |: |||
```

```
Db      173 DWPADRLNVYILDD-  ---- --    -  - - ---------------- ------ 187

Qy      341 DKVQPSFVKERRAMKREYEEYKVRVNALVAKAQKTPEEGWTMQDGTPWPGNNTRDHPGHD 400
                        :: : :: |  |
Db      188 -- ----------RPKAPHDFAVEAGA-------------------- --------- -- 231

Qy      401 SGLPWEILGARDIPFNELPRLVYVSREKRPGYQHHKKAGAENALVRVSAVLTNAPYILNV 460
                                  |: | :    :| :||  |  : |    |::: :
Db      232 -----    -       ----GYIIPAE -- NNHAKAGNLNHALAV-  -TDSPFAVIF 282

Qy      461 DCDHYVNNSKAVREAMCILMDPQVGRDVCYVQFPQRFDGIDKSDRYANRNV------VF 513
            ||||        |    :: ||    ::  :| |  |   |   |     :       :!
Db      283 DCDHVPTRGFLRRTGWMMADP   -NLALLQTPHHFYAPDPFQRNLAGGMHVPPEGNMF 286

Qy      514 FDVNMKGLDGIQGPVYVGTGCVFNRQALYGYGPPSMPSLRKRKDSSSCFSCCCPSKKKPA 573
            : :  |      : |:  :  ::|: : |
Db      289 YGLVQPGNDFWDAT FGGSCAIIRREAVMGIG   - ---------------- ------ 320

Qy      574 CDPAEVYEPAKEP NAAIEN.TRIFMYCFHFKSMLISQLSFEKTFGLSSVFIESTLMEN 633

Db      321  -                                          - ---------   - -- 320

Qy      634 GGVPESANSPFFIK AIQVIGGGYEEKTEWGKQIGWIYGSVTEDILSGFKMHCRGWRSIY 693
                                                      |:   :||||  :  ||  ||| : |
Db      321 -------------- ----    - ---------GFATETVTEDAHTALKMQRRGWGTAY 346

Qy      694 CMPVRPAFKGSAPINLSDRLHQVLRWALGSVFIFFSRHCPLWYGFGGGRLKWLQRLAYIN 753
                :  :|     :  : :  :||  | ::'    :  |     |  |:  |||  ':
Db      347 LR- -EPLAAGLATEH.ILHIGQRVRWARGMIQIMRLDNPML-----GAGLRWFQRLCYLS 399

Qy      754 TIVYPFTSLPLIAY'TIPAVCLLTGKFIIPTLSNLASMLFL--GLFISIIVTAVLELRWS 811
                :  :   ::| : :   |      |: :|   | ||  ::    :| ':|  : :| ||
Db      400 AMSHFLFAIPRLTF.VSPLAFLFLGQNII AASPLAISVYALPHIFHSVITLSRIEGRW- 457

Qy      812 GVSIEDLWRNEQFW VIGGVSAHLFAVFQGFLKMLAGIDTNFTVTAKAAEDAEFGELYMV 870
                    ||  :   :  ||  ,   .:|      | || |
Db      458 --- -- --RYSFW-EIYETSLALFLVRITIVTLLQPHKGKFNVTDKGG  --------- 497

Qy      871 KWTTLLIPPTTLLIINMSGCAGFSDALNKGYEAWGPLFGKVFFAFWVILHLYPFLKGLMG 930
                                  : :||   '   ::  |  |
Db      498 ------------ --  ---------DLARGYFDWDAVYPNVILA---   -- ------  516

Qy      931 RQNLTPTIVVLWSVLLASVFSLVW 954
                  || : ::  || ::|
Db      517 -- --   GVLCAALLRGVFGIVW 532


RESULT 10
BCA3_ACEXY
ID   BCA3_ACEXY     STANDARD;      PRT;   745 AA.
AC   Q9WX61;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Cellulose synthase 1 catalytic subunit [UDP forming] (EC 2.4.1.12)
GN   BCSA1.
OS   Acetobacter xylinus.
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;
OC   Acetobacteraceae; Gluconacetobacter.
OX   NCBI_TaxID=28448;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JCM 7664 / IFO 13693;
RX   MEDLINE=99310341; PubMed=10382968;
RA   Umeda Y., Hirano A., Ishibashi M., Akiyama H., Onizuka T., Ikeuchi M.,
RA   Inoue Y.;
RT   "Cloning of cellulose synthase genes from Acetobacter xylinum JCM
RT   7664: implication of a novel set of cellulose synthase genes.";
RL   DNA Res. 6:109-115(1999).
CC   -!- FUNCTION: Catalytic subunit of cellulose synthase. It polymerizes
CC       uridine 5'-diphosphate glucose to cellulose. The thick cellulosic
CC       mats generated by this enzyme probably provide a specialized
CC       protective environment to the bacterium (By similarity).
```

```
CC   -!- CATALYTIC ACTIVITY: UDP-glucose + {(1,4)-beta-D-glucosyl}(N) = UDP
CC       + {(1,4)-beta-D-glucosyl}(N+1).
CC   -!  COFACTOR: Magnesium (By similarity).
CC   -!- ENZYME REGULATION: Activated by bis-(3'-5') cyclic diguanylic acid
CC       (c-di-GMP) (By similarity).
CC   -!- PATHWAY: Bacterial cellulose biosynthesis.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC       (Potential).
CC   -!- DOMAIN: There are two conserved domains in the globular part of
CC       the catalytic subunit: the N-terminal domain (domain A) contains
CC       the conserved DXD motif and is possibly involved in catalysis and
CC       substrate binding. The C-terminal domain (domain B) contains the
CC       QXXRW motif and is present only in processive glycosyl
CC       transferases. It could be involved in the processivity function of
CC       the enzyme, possibly required for holding the growing glycan chain
CC       in the active site.
CC    !- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB015802; BAA77585.1; -.
DR   InterPro; IPR003919; Cellsynth_A.
DR   InterPro; IPR001173; Glyco_trans_2.
DR   Pfam; PF00535; Glycos_transf_2; 1.
DR   PRINTS; PR01439; CELLSNTHASEA.
KW   Cellulose biosynthesis; Transferase; Glycosyltransferase;
KW   Transmembrane; Inner membrane.
FT   DOMAIN       147     240       CATALYTIC SUBDOMAIN A
FT   DOMAIN       317     377       CATALYTIC SUBDOMAIN B.
FT   TRANSMEM      29      49       POTENTIAL.
FT   TRANSMEM     106     126       POTENTIAL.
FT   TRANSMEM     153     173       POTENTIAL.
FT   TRANSMEM     407     427       POTENTIAL.
FT   TRANSMEM     430     450       POTENTIAL.
FT   TRANSMEM     468     488       POTENTIAL.
FT   TRANSMEM     515     535       POTENTIAL.
FT   TRANSMEM     547     567       POTENTIAL.
FT   TRANSMEM     649     669       POTENTIAL.
FT   ACT SITE     189     189       POTENTIAL.
FT   ACT SITE     333     333       POTENTIAL.
FT   BINDING      236     236       SUBSTRATE (POTENTIAL)
FT   BINDING      238     238       SUBSTRATE (POTENTIAL)
SQ   SEQUENCE   745 AA;  83518 MW;  5[illegible]
```

Query Match 4.0%; Score 205; DB 1; Length 745;
Best Local Similarity 20.3%; Pred. No. 1.7e-07;
Matches 143; Conservative 64; Mismatches 204; Indels 294; Gaps 27;

```
Qy      177 IIMRLVILGLFFHFRITNPVDSAFGLW------LTSVICEIWFAFSWVLDQFPKWNPVNR 230
Db       80 VLSALVSL-RYLTWRLTETLD--FDTWTQGILGVTLLLAELYALYMLFLSYFQTISPLHR 136

Qy      231 ETYIERLSARYEREGEPSQLAGVDFFVSTVDPLKEPPLITANTVLSILAVDYPVDKVSCY 290
Db      137 --------APLPLPANPDEWPTVDIFIPTYD---EALSIVRLTVLGALGIDWPPDKVNVY 185

Qy      291 VSDDGAAMLSFESLVETAEFARKWVPFCKKFSIEPRAPEFYFSQKIDYLKDKVQPSFVKE 350
Db      186 ILDDG----------RREEFAR------------------FA---------------- 199

Qy      351 RRAMKRDYEEYKVRVNALVAKAQKTPEEGWTMQDGTPWPGNNTRDHPGHDSGLPWEILGA 410
Db      200 ------------------------------------------------EACGA 204

Qy      411 RDIEGNELPRLVYVSREKRPGYQHHKKAGAENALVRVSAVLTNAPYILNVDCDHYVNNSK 470
Db      205 R--- -------YIARPDNA----HAKAGNLNYAIK----HTGDHILILDCDHI----- 240

Qy      471 AVREAMCILMDPQVG-RDVCYVQFPQRFDGIDKSDRYANRNVVFFDVNMKGLDGIQGPVY 529
Db      241 PTRAFLQISMGWMVSDSNIALLQTPHHFYSPDP-------------------------- 273

Qy      530 VGTGCVFNRQALYGYGPPSMPSLRKRKDSSSCFSCCCPSKKKPAQDPAEVYRDAKREDLN 589
Db      274 ------FQRNLAVGYRTPPEGNL----------------------------------- 290

Qy      590 AAIFNLTEIDNYDEHERSMLISQLSFEKTFGLSSVFIESTLMENGGVPESANSPPFIKEA 649
Db      291 --FYGVIQ-DGND----------FWDATFFCGSCAI-----------------LRRKA 318

Qy      650 IQVIGCGYEEKTEWGKQIGWIYGSVTEDILSGFKMHCRGWRSIYCMPVRPAFKGSAPINL 709
Db      319 IEEIG-----------GFATETVTEDAHTALRMQRKGWSTAYLR--IPLASGLATERL 363

Qy      710 SDRLHQVLRWALGSVEIFFSRHCPLWYGFGGGRLKWLQRLAYINTIVYPFTSLPLIAYCT 769
Db      364 ITHIGQRMRWARGMIQIFRVDNPML-----GSGLKLGQRLCYLSAMTSFFFAIPRVIFLA 418

Qy      770 IPAVCLLTGKFIIPTLSNLASMLFLG------LFISIIVTAVLELRWSGVSIEDLWRNEQ 823
Db      419 SPLAFLFFSQNII-----AASPLAVGVYAIPHMFHSIATAAKVNKGW----------RYS 463

Qy      824 FW--VIGGVSAHLFAVFQGFLKMLAGIDTNFTVTAKAA--EDAEF 864
Db      464 FWSEVYETVMA-LFLVRVTIVTMLFPSKGKFNVTEKGGVLEREEF 507
```

```
RESULT 11
ACS2_ACEXY
ID   ACS2_ACEXY     STANDARD;      PRT;  1596 AA.
AC   Q59167;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Cellulose synthase 2 [Includes: Cellulose synthase catalytic subunit
DE   [UDP-forming] (EC 2.4.1.12); Cyclic di-GMP binding domain (Cellulose
DE   synthase 2 regulatory domain)].
GN   ACSAII.
OS   Acetobacter xylinus.
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;
OC   Acetobacteraceae; Gluconacetobacter.
OX   NCBI_TaxID=28448;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 23769;
RX   MEDLINE=95394846; PubMed=7665515;
RA   Saxena I.M., Brown R.M. Jr.;
RT   "Identification of a second cellulose synthase gene (acsAII) in
RT   Acetobacter xylinum.";
RL   J. Bacteriol. 177:5276-5283(1995).
CC   -!- CATALYTIC ACTIVITY: UDP-glucose + {(1,4)-beta-D-glucosyl}(N) = UDP
CC       + {(1,4)-beta-D-glucosyl}(N+1).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC       (Potential).
CC   -!- DOMAIN: There are two conserved domains in the globular part of
CC       the catalytic subunit: the N-terminal domain (domain A) contains
CC       the conserved DXD motif and is possibly involved in catalysis and
CC       substrate binding. The C-terminal domain (domain B) contains the
CC       QXXRW motif and is present only in processive glycosyl
CC       transferases. It could be involved in the processivity function of
CC       the enzyme, possibly required for holding the growing glycan chain
CC       in the active site.
CC   -!- MISCELLANEOUS: It is not essential for cellulose production in
CC       this strain.
CC   -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE
CC       GLYCOSYLTRANSFERASE FAMILY 2.
CC   -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE ACSB/BCSB
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
C   between the Swiss Institute of Bioinformatics and the EMBL outstation -
C   the European Bioinformatics Institute. There are no restrictions on its
C   use by non-profit institutions as long as its content is in no way
C   modified and this statement is not removed. Usage by and for commercial
C   entities requires a license agreement (See http://www.isb-sib.ch/announce/
C   or send an email to license@isb-sib.ch).
C   ----------------------------------------------------------------------
R   EMBL; U15957; AAA85264.1; -.
R   PIR; T31338; T31338.
R   InterPro; IPR003919; Cellsynth_A.
R   InterPro; IPR003920; Cellsynth_B.
R   InterPro; IPR001173; Glyco_trans_2.
R   Pfam; PF03170; BcsB; 1.
R   Pfam; PF00535; Glycos_transf_2; 1.
R   PRINTS; PR01439; CELLSNTHASEA.
R   PRINTS; PR01440; CELLSNTHASEB.
W   Cellulose biosynthesis; Transferase; Glycosyltransferase;
W   Transmembrane; Inner membrane.
T   DOMAIN        1    749       CATALYTIC.
T   DOMAIN      750   1596       CYCLIC DI-GMP BINDING (POTENTIAL).
T   DOMAIN      145    238       CATALYTIC SUBDOMAIN A.
T   DOMAIN      315    375       CATALYTIC SUBDOMAIN B.
T   TRANSMEM     25     45       POTENTIAL.
T   TRANSMEM    106    126       POTENTIAL.
T   TRANSMEM    396    416       POTENTIAL.
T   TRANSMEM    421    441       POTENTIAL.
T   TRANSMEM    505    525       POTENTIAL.
T   TRANSMEM    544    564       POTENTIAL.
T   TRANSMEM   1553   1573       POTENTIAL.
T   ACT_SITE    187    187       POTENTIAL.
T   ACT_SITE    331    331       POTENTIAL.
T   BINDING     234    234       SUBSTRATE (POTENTIAL).
T   BINDING     236    236       SUBSTRATE (POTENTIAL).
SQ  SEQUENCE   1596 AA;  175799 MW;  D98A6F6259E1F3CE CRC64;

 Query Match             4.0%;  Score 207.5;  DB 1;  Length 1596;
 Best Local Similarity   17.8%;  Pred. No. 5.9e-07;
 Matches  140;  Conservative   87;  Mismatches  229;  Indels  331;  Gaps   25;

Qy      176 VIIMRLVILGLFFHFRITNPVDSAFGLWLTS-------VICEIWFAFSWVLDQFPKWNPV 228
            :: :| ::  |     : :|: :|  | | :        :|  :|  || ||: |   |:
Db       82 LVSLRYMVWRLTTTLELHSPLQAALSLLLVAAELYALLTLCLSYFQMSWPLDRKPL--PL 139

Qy      229 NRETYIERLSARYEREGEPSQLAGVDFFVSTVDPLKEPPLITANTVLSILAVDYPVDKVS 288
              :|              :    || :| : :   |   :  :|||  ||:|:| ||::
Db      140 PADT--------------TDWPVVDVYVPSYN---EELSLVRSTVLGALAIDWPADKLN 181

Qy      289 CYVSDDGAAMLSFESLVETAEFARKWVPFCKKFSIEPRAPEFYFSQKIDYLKDKVQPSFV 348
             |: |||                        |   |:             :|
Db      182 VYILDDG-----------------------------RRKSFH-------------AFA 197

Qy      349 KERRAMKRDYEEYKVRVNALVAKAQKTPEEGWTMQDGTPWPGNNTRDHPGHDSGLPWEIL 408
             |  |                           |: ::|
Db      198 MEAGA------------------------GYIIRD------------------------ 208

Qy      409 GARDIEGNELPRLVYVSREKRPGYQHHKKAGAENALVRVSAVLTNAPYILNVDCDHYVNN 468
                                    :| |||  |  :||    |   |::  ||||
Db      209 ----------------------QNNHAKAGNLNHALRV----TEGEYVVIFDCDHIPTR 241

Qy      469 SKAVREAMCILMDPQVGRDVCYVQFPQRFDGIDKSDR--YANRNV-----VFFDVNMKGL 521
              :   :: ||:   : :| |  |  |  |    :||     :|: :   |
Db      242 GFLKKTIGWMMADPK----LALLQTPHHFYSPDPFQRNLATGQNVPPEGNMFYGLVQDGN 297

Qy      522 DGIQGPVYVGTGCVFNRQALYGYGPPSMPSLRKRKDSSSCFSCCCPSKKKPAQDPAEVYR 581
             |      : |:     | |: | |
Db      298 DFWDATFFCGSCAAIRRSAVLGIG---------------------------------- 321

Qy      582 DAKREDLNAAIFNLTEIDNYDEHERSMLISQLSFEKTFGLSSVFIESTLMENGGVPESAN 641

Db      322 ------------------------------------------------------------ 321

Qy      642 SPPFIKEAIQVIGCGYEEKTEWGKQIGWIYGSVTEDILSGFKMHCRGWRSIYCMPVRPAF 701
                                    |:   :|||| :  ||   || : |    :|
```

```
Db      322 ---- ---------- -  ------- GFATETVTEDAHTALKMQREGWHTAYLR--CPLA 353

Qy      702 KGSAPINLSDRLHQVLRWALGSVEIFFSRHCPLWYGFGGGRLKWLQRLAYINTIVYPFTS 761
             | :   :   : : :||| | ::    : ||     |  |:| ||| |:: : :   :
Db      354 AGLSTERLMLHIGQRVRWARGMLQIMRLDN-PLL----GSGLRWQQRLCYLSAMSHFLFA 408

Qy      762 LPLIAYCTIPAVCLLTGKFIIPTLSNLASML  FLGLFISIIVTAVLELRWSGVSIEDLW 819
             :| : :   |   |  |: ||   |  : ::  :  :| ||   : :| ||
Db      409 IPRLVFLASPLAFLFLGQNII-AASPFAILVYAFFHVFHSIGTLSRVEGRW--------- 458

Qy      820 RNEQFW-VIGGVSAHLFAVFQGFLKMLAGIDTNFTVTAKAA-EDAEFGELYMVKWTTLLI 877
              ||  |   :  || |    : :|      | || |     :|: :|  |    :|
Db      459 -RYSFWSEIYETTLALFLVRVTIMTLLNPRKGEFNVTDKGGLLQSEYFDLNAVYPNVILA 517

Qy      878 PPTTLLIINMSGCAGF     -------SDALNKGYEAWGPLFGKVFFAFWVILHLYPFLK 926
              | ::   :  |              | |||                 || :  |   |
Db      518 VILALALVR--GIGQMMWEYHDRLAQSFALN---------------TLWVAVSLIIVLA 560

Qy      927 GL-MGRQ 932
             : :||:
Db      561 SIAVGRE 567
```

```
RESULT 12
BCSA_XANAC
ID   BCSA_XANAC     STANDARD;      PRT;   729 AA.
AC   P58932;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Cellulose synthase catalytic subunit [UDP-forming] (EC 2.4.1.12).
GN   BCSA OR XAC3518.
OS   Xanthomonas axonopodis (pv. citri).
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC   Xanthomonadaceae; Xanthomonas.
OX   NCBI_TaxID=92829;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=306 / ATCC 13902 / XV 101;
RX   MEDLINE=22022145; PubMed=12024217;
RA   da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA   Quaggio R.B., Monteiro Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA   Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA   Camarotte G., Cannavan F., Cardozo J., Chambergo F., Ciapina L.P.,
RA   Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,
RA   Faria J.B., Ferreira A.C.S., Ferreira R.C.C., Ferro M.I.T.,
RA   Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA   Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA   Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA   Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA   Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA   Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA   Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA   Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA   Setubal J.C., Kitajima J.P.;
RT   "Comparison of the genomes of two Xanthomonas pathogens with differing
RT   host specificities.";
RL   Nature 417:459-463(2002).
CC   -!- FUNCTION: Catalytic subunit of cellulose synthase. It polymerizes
CC       uridine 5'-diphosphate glucose to cellulose, which is produced as
CC       an extracellular component for mechanical and chemical protection
CC       (By similarity).
CC   -!- CATALYTIC ACTIVITY: UDP-glucose + {(1,4)-beta-D-glucosyl}(N) = UDP
CC       + {(1,4)-beta-D-glucosyl}(N+1).
CC   -!- COFACTOR: Magnesium (By similarity).
CC   -!- ENZYME REGULATION: Activated by bis-(3'-5') cyclic diguanylic acid
CC       (c-di-GMP) (By similarity).
CC   -!- PATHWAY: Bacterial cellulose biosynthesis.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC       (Potential).
CC   -!- DOMAIN: There are two conserved domains in the globular part of
CC       the protein: the N-terminal domain (domain A) contains
CC       the conserved DXD motif and is possibly involved in catalysis and
```

```
CC       substrate binding. The C-terminal domain (domain B) contains the
CC       QXXRW motif and is present only in processive glycosyl
CC       transferases. It could be involved in the processivity function of
CC       the enzyme, possibly required for holding the growing glycan chain
CC       in the active site.
CC   -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
CC   ---------------------------------------- ------- ----- ----------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---- ---- ---- ---------------- ---- ----- ----------------------
DR   EMBL; AE012000; AAM38361.1; -.
DR   InterPro; IPR001173; Glyco_trans 2.
DR   Pfam; PF00535; Glycos_transf_2; 1.
KW   Cellulose biosynthesis; Transferase; Glycosyltransferase;
KW   Transmembrane; Inner membrane; Complete proteome.
FT   DOMAIN      151    244       CATALYTIC SUBDOMAIN A.
FT   DOMAIN      321    381       CATALYTIC SUBDOMAIN B.
FT   TRANSMEM     30     50       POTENTIAL.
FT   TRANSMEM    110    130       POTENTIAL.
FT   TRANSMEM    171    191       POTENTIAL.
FT   TRANSMEM    405    425       POTENTIAL.
FT   TRANSMEM    427    447       POTENTIAL.
FT   TRANSMEM    520    540       POTENTIAL.
FT   TRANSMEM    549    569       POTENTIAL.
FT   TRANSMEM    610    630       POTENTIAL.
FT   ACT_SITE    193    193       POTENTIAL.
FT   ACT_SITE    337    337       POTENTIAL.
FT   BINDING     240    240       SUBSTRATE (POTENTIAL).
FT   BINDING     242    242       SUBSTRATE (POTENTIAL).
SQ   SEQUENCE   729 AA;  80915 MW;  B9C08BB995E795B1 CR64;

  Query Match             3.5%;  Score 181;  DB 1;  Length 729;
  Best Local Similarity   18.1%;  Pred. No. 1.7e-05;
  Matches  137;  Conservative   78;  Mismatches  199;  Indels  342;  Gaps   27;

Qy        174 RAVII---MRLVILGLFFHFRIT-----NPVDSAFGLWLLS-------VICEIWFAFS 216
              | |||:   | | :   :  :|:|     : ||   ||          ::   :|
Db         76 RVVILMMMGMSLAVSCRYIWWRMTQTMGVGSAVDFILGLGL-GAELYAFVILVLGYFQVL 135

Qy        217 WVLDQFPKWNPVNRETYIERLSAPVERSGEPSQLAGVDFFVSTVDPLKEPPLITANTVLS 276
              | |:: |   | ::  :              :|    ||    | :   ||  :   :||:
Db        136 WPLNRKPVPLPADQRLW----          -PS- - VDVP- TYN---RPLSVVRTTVLA 175

Qy        277 ILAVDYPVDKVSCYVSDDGAAMLSFEALVETAEFARRWLPFCRKYSIEPRAPEFYFSQKI 336
                 :|:|  |:: :: |||                                 |
Db        176 ASVIDWPAGKITIHLLDDG---          PVGEFF    A- APVQIN ---------   209

Qy        337 DYLKDKVQPSFVKERRAMKRDYEEYKVRVNALVAKAQKIPE-GWTMQDGTPWPGNNTRDH 396

Db        210 -------------- ----- ------- ------ ----    ---------------- 209

Qy        397 PGHDSGLPWEILGARDIEGNELPRLVYVSREKRPGYQHHKK-GAENALVRVSAVLTNAPY 456
                                  ||:|        : |:-  || ::      :  |
Db        210 ----------------------- YVTPTNNA----HAF-GN1NAALK----KCSGDY 235

Qy        457 ILNVDCDHYVNNSKAVREAMCILMDPQVGRDVCYVQFPQRF ----  DGIDKSDRYANR 509
              :  ||||    |         | | :    . |, |           :|   :  |
Db        236 VAIFDCDHIPTRSFLQVAMGWFLHDTK- -LALVQMPHYF SPDPFERNLDTHGKVPNE 291

Qy        510 NVVFFDVNMKGLDGIQGPVYVGTGCVFNRQALYGYGPPSMP-LRKRKDSSSCFSCCCPSK 569
              :|: :   | |     : |:  ' | |
Db        292 GELFYGLIQDGNDQWNATFFCGSCAVIKRTA - -----   ------------------ 322

Qy        570 KKPAQDPAEVYRDAKREDLNAAIFNLTEILNYDEHERSML QLSFEKTFGLSSVFIEST 629

Db        323 ---------------------- -  -- -- ----     ------- ------- 322

Qy        630 LMENGGVPESANSPFFIKEAIQVIGCGYEEKTEWGKQIGW GSVTEDILSGFKMHCRGW 689

              | | |||           |::                    :|||| : |: ||:
Db        323 LEEVGGV-----------AVE--------------------TVTEDAHTALKLQRRGY 349

Qy        690 RSIYCMPVRPAFKGSAPINLSDRLHQVLRWALGSVEIFFSRHCPLWYGFGGGRLKWLQRL 749
              |: |     |  | |  :||  : | :||| |  :|    : ||     |  ||  |||
Db        350 RTAYL--AVPQAAGLATESLSGHVAQRIRWARGMAQIARIDN-PLL----GRGLKLSQRL 402

Qy        750 AYINTIVYPFTSLPLIAYCTIPAVCLLTGKFIIPTLSNLASMLFLGLFISIIVTAVLELR 809
                |:| ::: |  :| | | | |   |  |  :|      || |   : ::  :  :|:
Db        403 CYLNAMLHFFYGVPRIIYLTAPLAYLFFGAHVIQ-----ASAL---MILAYALPHILQAN 454

Qy        810 WSGVSIEDLWR----NEQF-----WVIGGVSAHLFAVFQGFLKMLAGIDT-NFTVT---- 855
                : : ::  :|    || :     | |          |:  | |       | ||
Db        455 LTNLRVQSRFRHLLWNEVYETTLAWYI----------FRPTLVALLNPKLGKFNVTPKGG 504

Qy        856 --AKAAEDAEFGELYMVKWTTLLIPPTTLLIINMSG 889
                |::  ||:  : |:          ||::|: |
Db        505 LVARSYFDAQIAKPYLF----------LLLLNVVG 529


RESULT 13
ACS1_ACEXY
ID   ACS1_ACEXY     STANDARD;      PRT;  1550 AA.
AC   P21877; P37717;
DT   01-MAY-1991 (Rel. 18, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Cellulose synthase 1 [Includes: Cellulose synthase catalytic domain
DE   [UDP-forming] (EC 2.4.1.12); Cyclic di-GMP binding domain (Cellulose
DE   synthase 1 regulatory domain)].
GN   ACSAB OR ACSA OR ACSB.
OS   Acetobacter xylinus.
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;
OC   Acetobacteraceae; Gluconacetobacter.
OX   NCBI_TaxID=28448;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 6-20.
RC   STRAIN=ATCC 53582;
RX   MEDLINE=91346705; PubMed=2151718;
RA   Saxena I.M., Lin F.C., Brown R.M. Jr.;
RT   "Cloning and sequencing of the cellulose synthase catalytic subunit
RT   gene of Acetobacter xylinum.";
RL   Plant Mol. Biol. 15:673-683(1990).
RN   [2]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 768-781.
RC   STRAIN=ATCC 53582;
RX   MEDLINE=91322509; PubMed=1830823;
RA   Saxena I.M., Lin F.C., Brown R.M. Jr.;
RT   "Identification of a new gene in an operon for cellulose biosynthesis
RT   in Acetobacter xylinum.";
RL   Plant Mol. Biol. 16:947-954(1991).
RN   [3]
RP   REVISIONS.
RC   STRAIN=ATCC 53582;
RX   MEDLINE=94364954; PubMed=8083166;
RA   Saxena I.M., Kudlicka K., Okuda K., Brown R.M. Jr.;
RT   "Characterization of genes in the cellulose-synthesizing operon (acs
RT   operon) of Acetobacter xylinum: implications for cellulose
RT   crystallization.";
RL   J. Bacteriol. 176:5735-5752(1994).
RN   [4]
RP   SEQUENCE OF 1-8 FROM N.A.
RC   STRAIN=ATCC 23769;
RX   MEDLINE=94131945; PubMed=8300521;
RA   Standal R., Iversen T.-G., Coucheron D.H., Fjaervik E., Blatny J.M.,
RA   Valla S.;
RT   "A new gene required for cellulose production and a gene encoding
RT   cellulolytic activity in Acetobacter xylinum are colocalized with the
RT   bcs operon.";
RL   J. Bacteriol. 176:665-672(1994).
RN   [5]
RP   FUNCTION.
RC   STRAIN=ATCC 53582;
```

```
X   MEDLINE=90202941; PubMed=2138620;
A   Lin F.C., Brown R.M. Jr., Drake R.R. Jr., Haley B.E.;
T   "Identification of the uridine 5'-diphosphoglucose (UDP-Glc) binding
T   subunit of cellulose synthase in Acetobacter xylinum using the
T   photoaffinity probe 5-azido-UDP-Glc.";
L   J. Biol. Chem. 265:4782-4784(1990).
N   [6]
P   3D-STRUCTURE MODELING, AND MUTAGENESIS OF ASP-188; ASP-189; ASP-236;
P   ASP-333; GLN-369; ARG-370 AND ARG-372.
C   STRAIN=ATCC 23769;
X   MEDLINE=21324818; PubMed=11430986;
A   Saxena I.M., Brown R.M. Jr., Dandekar T.;
T   "Structure-function characterization of cellulose synthase:
T   relationship to other glycosyltransferases.";
L   Phytochemistry 57:1135-1148(2001).
N   [7]
P   REVIEW ON DOMAIN ARCHITECTURE.
X   MEDLINE=95189716; PubMed=7883697;
A   Saxena I.M., Brown R.M. Jr., Fevre M., Geremia R.A., Henrissat B.;
T   "Multidomain architecture of beta-glycosyl transferases: implications
T   for mechanism of action.";
L   J. Bacteriol. 177:1419-1424(1995).
C   -!- FUNCTION: Bifunctional protein comprised of a catalytic subunit
C       and a regulatory subunit. The catalytic subunit of cellulose
C       synthase polymerizes uridine 5'-diphosphate glucose to cellulose
C       in a processive way. The thick cellulosic mats generated by this
C       enzyme probably provide a specialized protective environment to
C       the bacterium. The regulatory subunit binds bis-(3'-5') cyclic
C       diguanylic acid (c-di-GMP).
C   -!- CATALYTIC ACTIVITY: UDP-glucose + {(1,4)-beta-D-glucosyl}(N) = UDP
C       + {(1,4)-beta-D-glucosyl}(N+1).
C   -!- COFACTOR: Magnesium (By similarity).
C   -!- ENZYME REGULATION: Activated by c-di-GMP.
C   -!- PATHWAY: Bacterial cellulose biosynthesis.
C   -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
C       (Potential).
C   -!- INDUCTION: Cellulose is produced at a linear rate with respect to
C       cell growth when O(2) is present.
C   -!- DOMAIN: There are two conserved domains in the globular part of
C       the catalytic subunit: the N-terminal domain (domain A) contains
C       the conserved DXD motif and is possibly involved in catalysis and
C       substrate binding. The C-terminal domain (domain B) contains the
C       QXXRW motif and is present only in processive glycosyl
C       transferases. It could be involved in the processivity function of
C       the enzyme, possibly required for holding the growing glycan chain
C       in the active site.
C    !- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE
C       GLYCOSYLTRANSFERASE FAMILY 2.
C   -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE ACSB/BCSB
C       FAMILY.
C   -!- CAUTION: Was originally (Ref.1) thought to be two separate ORFs
C       named acsA and acsB, due to a frameshift in position 678.
C   ---------------------------------------------------------------------
C   This SWISS-PROT entry is copyright. It is produced through a collaboration
C   between the Swiss Institute of Bioinformatics and the EMBL outstation -
C   the European Bioinformatics Institute. There are no restrictions on its
C   use by non-profit institutions as long as its content is in no way
C   modified and this statement is not removed. Usage by and for commercial
C   entities requires a license agreement (See http://www.isb-sib.ch/announce/
C   or send an email to license@isb-sib.ch).
C   ---------------------------------------------------------------------
R   EMBL; X54676; CAA38487.1; ALT_FRAME.
R   EMBL; X54676; CAA38488.1; ALT_FRAME.
R   EMBL; M96060; AAA16971.1; -.
R   InterPro; IPR003919; Cellsynth_A.
R   InterPro; IPR001173; Glyco_trans_2.
R   Pfam; PF03170; BcsB; 1.
R   Pfam; PF00535; Glycos_transf_2; 1.
R   PRINTS; PR01439; CELLSNTHASEA.
W   Cellulose biosynthesis; Transferase; Glycosyltransferase;
W   Transmembrane; Inner membrane.
T   DOMAIN       1    741       CATALYTIC.
T   DOMAIN     742   1550       CYCLIC DI-GMP BINDING.
FT  DOMAIN     147    243       CATALYTIC SUBDOMAIN A.
FT  DOMAIN     317    377       CATALYTIC SUBDOMAIN B.
FT  TRANSMEM    26     46       POTENTIAL.
FT  TRANSMEM    47     67       POTENTIAL.
FT  TRANSMEM   106    126       POTENTIAL.
FT  TRANSMEM   398    418       POTENTIAL.
FT  TRANSMEM   423    443       POTENTIAL.
FT  TRANSMEM   468    488       POTENTIAL.
FT  TRANSMEM   507    527       POTENTIAL.
FT  TRANSMEM   547    567       POTENTIAL.
FT  TRANSMEM  1513   1533       POTENTIAL.
FT  ACT_SITE   189    189       POTENTIAL.
FT  ACT_SITE   333    333       POTENTIAL.
FT  BINDING    236    236       SUBSTRATE (POTENTIAL).
FT  BINDING    238    238       SUBSTRATE (POTENTIAL).
FT  MUTAGEN    188    188       D->N: DECREASE IN ACTIVITY.
FT  MUTAGEN    188    188       D->E: LOSS OF ACTIVITY.
FT  MUTAGEN    189    189       D->Y: LOSS OF ACTIVITY.
FT  MUTAGEN    236    236       D->Y: LOSS OF ACTIVITY.
FT  MUTAGEN    333    333       D->E: LOSS OF ACTIVITY.
FT  MUTAGEN    369    369       Q->M: LOSS OF ACTIVITY.
FT  MUTAGEN    370    370       R->K: LOSS OF ACTIVITY.
FT  MUTAGEN    370    370       R->Q: DECREASE IN ACTIVITY.
FT  MUTAGEN    372    372       R->A: LOSS OF ACTIVITY.
SQ  SEQUENCE  1550 AA;  1[illegible]61 MW;  65AB6952BC39E961 CRC64;

  Query Match             [illegible]%;  Score 174.5;  DB 1;  Length 1550;
  Best Local Similarity   2[illegible].0%;  Pred. No. 0.00016;
  Matches  163;  Conservative   84;  Mismatches  253;  Indels  315;  Gaps   36;

Qy      141 VPTEQQMEEKPSAE[illegible]SEPLSIVYPIPRNKLTPYRAVIIMRLVILGL- --FFHFRITNPV 196
            :|.:|:        :    ::: ,  :|    |   |    |: ||    :  :|:|  :
Db       47 LPPEQQL  --- [illegible]AFVCVVIFFIVGHK- FSRRSQIFLEVLSGLVSLRYLTWRLTETL 98

Qy      197 DSAFGLWLTS -    VICEIWFAFSWVLDQFPKWNPVNRETYIERLSARYERDGEPSQL 250
            :  |'         :: |::       |  |     ::|      :       | :
Db       99 --SFDTWLQGLLGT[illegible]LLVAELYALMMLFLSYFQTIAFLHR--------APLPLPPNPDEW 148

Qy      251 AGVDFFVSTVDPLK[illegible]PPLITANTVLSILAVDYPVDKVSCYVSDDGAAMLSFESLVETAEF 310
              || |: :   .  :   |||  | :|:| ::|  :: |||
Db      149 PTVDIFVPTYN-   [illegible]EDSIVRLTVLGSLGIDWPPEKVRVHILDDG-- ------ ------ 190

Qy      311 ARKWVPFCKKFSIE[illegible]PAFEFYFSQKIDYLKDKVQPSFVKERRAMKRDYEEYKVRVNALVA 370
                        | |:|                :|  !
Db      191 -- ----------   RPFEF  -----   - --AAFAAE------- -----  ----- 201

Qy      371 KAQKTPEEGWTMQD[illegible]TPWPGNNTRDHPGHDSGLPWEILGARDIEGNELPRLVYVSREKRP 430
                                                          ||           |::|
Db      202 -------- ----      -  ---  ----------CGAN----------YIARPT   211

Qy      431 GYQHRKKAGAENAL[illegible]RVSAVLTNAPYILNVDCDHYVNNSKAVREAMCILMDPQVGRDVCY 490
              | :||  :          :  |:|  |''|    :       :: ||:   :
Db      212 --NEHAKAGNLNYA [illegible]---HTDGDYILIFDCDHVPTRAFLQLTMGWMVEDPK  --IAL 261

Qy      491 VQFPQRFDGIDKSD[illegible]YANRNVVFFDVNMKGLDGIQGPVYVGTGCVFNRQALYGYGPPSMP 550
            :| |  |                                                |:    ||  |
Db      262 MQTPHHFYSPDP-- ---------   --    ----------FQRNLSAGYRTPPEG 288

Qy      551 SLRKRKDSSSCFSC[illegible]TPSKKKPAQDFAEVYRDAKREDLNAAIFNLTEIDNYDEHERSMLI 610
            :|                                          :  :  | |
Db      289 NL---  -  --      -------  -  -  ------ ----FYGVVQ DGND- ------ 300

Qy      611 SQLSFEKTFGLSSV[illegible]FSTLMENGGVPESANSPFFIKEAIQVIGCGYEEKTEWGKQIGWI 670
              :: :    :  |                          : ||: ||          :
Db      301 ---FWDATFFCGSTAI---------- ---- --LRRTAIEQIG-------  ----GFA 326

Qy      671 YGSVTEDILSGFKM[illegible]RGWRSIYCMPVRPAFKGSAPINLSDRLHQVLRWALGSVEIFFSR 730
            :||||  :  :|  :| : :     |   | |   |  : | :|||  | :::|
Db      327 TQTVTEDAHTALKM[illegible]RLGWSTAYLR--IPLAGGLATERLILHIGQRVRWARGMLQIFRID 384

Qy      731 HCPLWYGFGGGRLKWLQRLAYINTIVYPFTSLPLIAYCTIPAVCLLTGKFIIPTLSNLAS 790
            : :|   || | | | |||  |:: :     ::| : : :|   |  |: |       ||
Db      385 N-PL---FGRG-LSWGQRLCYLSAMTSFLFAVPRVIFLSSPLAFLFFGQNII---- AAS 434
```

```
Qy        791 MLFLGLFISIIVTAVLELRWSGVSIEDLWRNEQFW-VIGGVSAHLFAVFQGFLKMLAGID 849
              | |   ::  :  :     :  |  ||  ||  :   :  || |    : :|:
Db        435 PLAL---LAYAIPHMFHAVGTASKINKGWR-YSFWSEVYETCMALFLVRVTIVTLLSPSR 490

Qy        850 TNFTVTAKAA--EDAEF--GELYMVKWTTLLIPPTTLLIINMSGCAGFSDALNKG-YE-A 903
              | || |    |   |  | :|          | :| : | |       | :| || :
Db        491 GKFNVTDKGGLLEKGYFDLGAVY-----   --PNIILGLIMFG------GLARGVYELS 534

Qy        904 WGPLFGKVFFAF-----WVILHLYPFLKGL-MGRQ 932
              :| |     |:     | :| |  |  : :|:
Db        535 FGHLDQIAERAYLLNSAWAMLSLIILAAIAVGRE 569


RESULT 14
A11B_HUMAN
ID   A11B_HUMAN     STANDARD;      PRT;  1177 AA.
AC   Q9Y2G3; Q96FN1; Q9UKK7;
DT   30-MAY-2000 (Rel. 39, Created)
DT   15-SEP-2003 (Rel. 42, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Potential phospholipid-transporting ATPase IF (EC 3.6.3.1) (ATPase
DE   class I type 11B) (ATPase IR).
GN   ATP11B OR ATPIF OR ATPIR OR KIAA0956.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE OF 1-248 FROM N.A.
RC   TISSUE=Colon;
RX   MEDLINE=20473714; PubMed=11015572;
RA   Halleck M.S., Lawler J.F. Jr., Blackshaw S., Gao L., Nagarajan P.,
RA   Hacker C., Pyle S., Newman J.T., Nakanishi Y., Ando H., Weinstock D.,
RA   Williamson P.L., Schlegel R.A.;
RT   "Differential expression of putative transbilayer amphipath
RT   transporters.";
RL   Physiol. Genomics 1:139-150(1999).
RN   [2]
RP   SEQUENCE OF 1-195 AND 540-1177 FROM N.A.
RC   TISSUE=Breast, and Colon;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [3]
RP   SEQUENCE OF 111-450 FROM N.A.
RX   MEDLINE=21895878; PubMed=11790799;
RA   Halleck M.S., Schlegel R.A., Williamson P.L.;
RT   "Reanalysis of ATP11B, a Type IV P-type ATPase.";
RL   J. Biol. Chem. 277:9736-9740(2002).
RN   [4]
RP   SEQUENCE OF 440-1111 FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=99246063; PubMed=10231032;
RA   Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirosawa M.,
RA   Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. XIII.
RT   The complete sequences of 100 new cDNA clones from brain which code
RT   for large proteins in vitro.";
RL   DNA Res. 6:63-70(1999).
RN   [5]
RP   SEQUENCE OF 688-1111 FROM N.A.
RC   TISSUE=Testis;
RA   Bloecker H., Boecher M., Brandt P., Mewes H.-W., Gassenhuber J.,
RA   Wiemann S.;
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + phosphate.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: Belongs to the cation transport ATPases family (P-type
CC       ATPases). Subfamily IV.
CC   -!- CAUTION: Ref.2 (AAH10630) sequence differs from that shown due to
CC       a frameshift in position 884.
CC   -!- CAUTION: Ref.2 (AAH42180) sequence differs from that shown due to
CC       the presence of C-terminal non-spliced intronic sequence.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF156548; AAF09446.1; -.
DR   EMBL; BC042180; AAH42180.1; ALT_TERM.
DR   EMBL; BC010630; AAH10630.1; ALT_FRAME.
DR   EMBL; AC069431; -; NOT_ANNOTATED_CDS.
DR   EMBL; AB023173; BAA76800.1; -.
DR   EMBL; AL133061; CAB61385.1; -.
DR   PIR; T42662; T42662.
DR   Genew; HGNC:13553; ATP11B.
DR   MIM; 605869; -.
DR   InterPro; IPR001757; ATPase_E1-E2.
DR   InterPro; IPR006539; Flippase.
DR   InterPro; IPR005834; Hydrolase.
DR   Pfam; PF00702; Hydrolase; 1.
DR   PRINTS; PR00119; CATATPASE.
DR   TIGRFAMs; TIGR01652; ATPase-Plipid; 1.
DR   TIGRFAMs; TIGR01494; ATPase_P-type; 4.
DR   PROSITE; PS00154; ATPASE_E1_E2; 1.
KW   Hydrolase; Transmembrane; Phosphorylation; Magnesium; Metal-binding;
KW   ATP-binding; Multigene family.
FT   DOMAIN       1     55       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    56     77       POTENTIAL.
FT   DOMAIN      78     82       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    83    104       POTENTIAL.
FT   DOMAIN     105    289       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   290    311       POTENTIAL.
FT   DOMAIN     312    341       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   342    359       POTENTIAL.
FT   DOMAIN     360    876       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   877    898       POTENTIAL.
FT   DOMAIN     899    910       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   911    930       POTENTIAL.
FT   DOMAIN     931    960       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   961    982       POTENTIAL.
FT   DOMAIN     983    997       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   998   1020       POTENTIAL.
FT   DOMAIN    1021   1025       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM  1026   1047       POTENTIAL.
FT   DOMAIN    1048   1065       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM  1066   1090       POTENTIAL.
FT   DOMAIN    1091   1177       CYTOPLASMIC (POTENTIAL).
FT   MOD_RES    407    407       PHOSPHORYLATION (BY SIMILARITY).
FT   METAL      821    821       MAGNESIUM (BY SIMILARITY).
FT   METAL      825    825       MAGNESIUM (BY SIMILARITY).
SQ   SEQUENCE   1177 AA;  134189 MW;  665110145457B220 CRC64;
```

```
Query Match              2.4%;  Score 127.5;  DB 1;  Length 1177;
Best Local Similarity   18.4%;  Pred. No. 0.35;
Matches  125;  Conservative  108;  Mismatches  242;  Indels  203;  Gaps   32;

y       445 VRVSAVLT----NAPYILNV---DCDHYVN--NSKAVREAMCILMDPQVGRDVCYVQFPQ 495
b       498 VQISNVQTDCTGDGPWQSNLAPSQLEYYASSPDEKALVEAAARIGIVFIGNS----EETM 553
y       496 RFDGIDKSDRYANRNVVFFDVNMKGLDGI-QGPVYVGTGCVFNRQALYGYGPPS--MP-- 550
b       554 EVKTLGKLERYKLLHILEFDSDRRRMSVIVQAP--------SGEKLLFAKGAESSILPKC 605
y       551 ---SLRKRKDSSSCF------SCCCPSKKKPAQDPAEVYRD--------AKREDLNAAIF 593
b       606 IGGEIEKTRIHVDEFALKGLRTLCIAYRKFTSKEYEEIDKRIFEARTALQQREEKLAAVF 665
y       594 NLTEIDNYDEHERSMLISQLSFEKTFGLSSVFIESTLMENGGVPESANSPPFIKEAIQV- 652
b       666 QFIEKD---------LIL---------LGATAVEDRLQDK---------- VRETIEAL 695
y       653 ----------------------IGCGYEEKT------- -----EWGKQIGWIYGSVTED 677
b       696 RMAGIKVWVLTGDKHETAVSVSLSCGHFHRTMNILELINQKSDSECAEQLRQLARRITED 755
y       678 -------ILSG-------------FKMHCRGWRSIYCMPVRPAFKGSA--PINLSDRLHQ 715
b       756 HVIQHGLVVDGTSLSLALREHEKLFMEVCRNCSAVLCCRMAPLQKAKVIRLIKISPEKPI 815
y       716 VLRWALGSVEIFFSRHCPLWYGFGG--------------GRLKWLQRLAYINTIVYPFTS 761
b       816 TLAVGDGANDVSMIQEAHVGIGIMGKEGRQAARNSDYAIARFKFLSKLLFVHGHFYYIRI 875
y       762 LPLIAYCTIPAVCLLTGKFI-----IPTLSNLASMLFLGLF------ISIIVTAVLE--- 807
b       876 ATLVQYFFYKNVCFITPQFLYQFYCLFSQQTLYDSVYLTLYNICFTSLPILIYSLLEQHV 935
y       808 ----LRWSGVSIEDLWRNE-------QFWVIGGVSAHLFAVFQGFLKMLAGIDTNFTVTA 856
b       936 DPHVLQNKPTLYRDISKNRLLSIKTFLYWTILGFS-HAFIFFFGSY-LLIGKDTSLLGNG 993
y       857 KAAEDAEFGELYMVKWTTLLIPPTTLLIINMSGCAGFSDALNKGYEAWGPLFGKVFFAFW 916
b       994 QMFGNWTFGTLV---FTVMVITVT----VKMALETHFWTWINH-LVTWGSIIFYFVFSLF 1045
y       917 VILHLYPFLKGLMGRQNLTPTIVVL------WSVLLASVFSLVWVKINPFVNKV------ 964
b      1046 YGGILWPFL----GSQNMYFVFIQLLSSGSAWFAIILMVVTCLFLDI---IKKVFDRHLH 1098
y       965 ----DNTLVAETCISIDC 978
b      1099 PTSTEKAQLTETNAGIKC 1116


ESULT 15
TCX_SCHPO
D   ATCX_SCHPO     STANDARD;      PRT;  1402 AA.
C   Q09891;
T   01-FEB-1996 (Rel. 33, Created)
T   01-FEB-1996 (Rel. 33, Last sequence update)
T   28-FEB-2003 (Rel. 41, Last annotation update)
E   Potential phospholipid transporting ATPase 1 (EC 3.6.3.1).
N   SPAC24B11.12C.
S   Schizosaccharomyces pombe (Fission yeast).
C   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
C   Schizosaccharomycetales; Schizosaccharomycetaceae;
C   Schizosaccharomyces.
X   NCBI_TaxID=4896;
N   [1]
P   SEQUENCE FROM N.A.
C   STRAIN=972;
X   MEDLINE=21848401; PubMed=11859360;
A   Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
A   Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA  Brocks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA  Collins M., Connor R., Corbin A., Davis P., Feltwell T., Fraser A.,
RA  Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA  Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA  James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA  Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA  Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,
RA  Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA  Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA  Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA  Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA  Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA  Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA  Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA  Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA  Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA  Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA  Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA  Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA  Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA  Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA  Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT  "The genome sequence of Schizosaccharomyces pombe.";
RL  Nature 415:871-880(2002).
CC  -!- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + phosphate.
CC  -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC  -!- SIMILARITY: Belongs to the cation transport ATPases family (P type
CC      ATPases). Subfamily IV.
CC  ---------- ---- - ------ --- - ---------- ------ --
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  ---------- ---- - ------ --- - ---------- ------ --
DR  EMBL; Z67757; CAA91777.1; -.
DR  PIR; T38339; S62557.
DR  GeneDB_SPombe; SPAC24B11.12c; .
DR  InterPro; IPR001757; ATPase_E1-E2.
DR  InterPro; IPR006539; Flippase.
DR  InterPro; IPR005834; Hydrolase.
DR  Pfam; PF00702; Hydrolase; 1.
DR  PRINTS; PR00119; CATATPASE.
DR  TIGRFAMs; TIGR01652; ATPase-Plipid; 1.
DR  TIGRFAMs; TIGR01494; ATPase_P-type; 8.
DR  PROSITE; PS00154; ATPASE_E1_E2; 1.
KW  Hydrolase; Transmembrane; Phosphorylation; Magnesium; ATP-binding.
FT  TRANSMEM     109     129       POTENTIAL.
FT  TRANSMEM     135     155       POTENTIAL.
FT  TRANSMEM     457     477       POTENTIAL.
FT  TRANSMEM     501     521       POTENTIAL.
FT  TRANSMEM    1066    1086       POTENTIAL.
FT  TRANSMEM    1101    1121       POTENTIAL.
FT  TRANSMEM    1151    1171       POTENTIAL.
FT  TRANSMEM    1193    1213       POTENTIAL.
FT  TRANSMEM    1218    1238       POTENTIAL.
FT  TRANSMEM    1260    1280       POTENTIAL.
FT  MOD_RES      569     569       PHOSPHORYLATION (BY SIMILARITY).
SQ  SEQUENCE   1402 AA;  15[illegible]359 MW;  E8495A7598B3020B CRC64;

  Query Match              2.2%;  Score 117;  DB 1;  Length 1402;
  Best Local Similarity   19.4%;  Pred. No. 2.7;
  Matches  131;  Conservative   96;  Mismatches  234;  Indels  214;  Gaps   34;

Qy        437 KAGAENALVRVSAVLTNA---PYILNVDCDHYVNNSKAVREAMCILMDPQVGRDVCYVQF 493
Db        660 KAGEEQSLACYEFFLALALCHSVVADRVGDRIVYKAQSPDEAALV----GTARDVGFVFL 715

Qy        494 PQRFD-----GIDKSDRYANRNVVFFDVNMKGLDGI-QGP----VYVGTGC---VFNRQA 540
Db        716 DQRRDIMVTRALGETQRFKLMDTIEFSSARKRMSVIVKGPDNRYVLICKGADSIIFER-- 773
```

```
Qy        541 LYGYGPPSMPSLRK------RKDSSSCFSCCCPSKKKPAQDPAEVYRDAKREDLNAAIFN 594
Db        774 ---LEPNEQVELRKTTSEHLRIFALEGLRTLCIAKRELTEE--EYYEWKEKYDIAA---- 824

Qy        595 LTEIDNYD---EHERSMLISQLSFEKTFGLSSVFIESTLMENGGVPES------------ 639
Db        825 -SAIENREEQIEEVADLIESHLTL-----LGGTAIEDRLQE-GVPDSIALLAQAGIKLW 876

Qy        640 -------- ------------ ------ ---ANSPPFIKEAIQVIGCGY-------- 657
Db        877 VLTGDKMETAINIGFSCNLLDAGMDMIKFDVDQEVSTP---  -ELEVILADYLYRYFGLS 931

Qy        658 ---- -- ---EEKTEWGKQIGWIYGSVTEDILSG-----FFMHCRGWRSIYCMPVRPAF 701
Db        932 GSVEELEAAKKDHDTPSGSHALVIDGSVLKRVLDGPMRTKFLLCKRCKAVLCCRVSPAQ 991

Qy        702 KGSAPINLSDRLHQVLRWAL--GSVEIFFSRHCPLWYGFGG ------------GRLKW 745
Db        992 KADV-VQLVRESLEVMTLAIGDGANDVAMIQKADIGVGIVGREGRAAAMSADYAIGQFRF 1050

Qy        746 LQRLAYIN---------------- TIVYPFTSLPLIAYGTIPAVCLLTGKFIIPTLSN 787
Db       1051 LSKLVLVHGRWDYNRVAEMVNNFFYKSVVWTFTLFWYQIYNNFDANYLFDYTYVM----- 1105

Qy        788 LASMLFLGLFISII------VTAVLELRWSGVSIEDLWRNEQF------ ----WVIGGV 830
Db       1106 LFNLIFSSLPVIVMGVYDQDVNADLSLR----TPQLYKRGILQLNSARKIFIGYMLDGF 1160

Qy        831 SAHLFAVFQGFLKMLAGIDTNFTVTAKAAEDA----EFGELYMVKWTTLLIPPTTLLIIN 886
Db       1161 YQSVICFFFSFL-----VINNVTTAAQNGRDTMAVQDLG-VYVAAPTIMVV--DTYVILN 1212

Qy        887 MSGCAGFSDALNKGYEAWGPLFGKVFFAFWV---  -ILHLYPFLKGLMGRQNLTPTIVVL 941
Db       1213 QSNWDVFSIGL------WA--LSCVTFWFWTGVYSQSLYTYEFYKS-ASRIFRTPN---F 1260

Qy        942 WSVLLASVFSLVWVK 956
Db       1261 WAVLCGTIVSCLFPK 1275
```

Search completed: October 29, 2003, 18:29:33
Job time : 31.8085 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 29, 2003, 18:25:58 ; Search time 68.1006 Seconds
(without alignments)
3785.922 Million cell updates/sec

Title: US 09-980-043-2
Perfect score: 5240
Sequence: 1 MMESGAPICHTCGEQVGHDA.....PFVNKVDLTLVARTCISIDC 978

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_23:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	5240	100.0	978	10	O81368	O81368 populus tre
2	4524	86.3	974	10	P93155	P93155 gossypium h
3	4426	84.5	974	10	Q8W1W0	Q8w1w0 gossypium h
4	4323.5	82.5	985	10	Q9C5Z8	Q9c5z8 arabidopsis
5	4316.5	82.4	985	10	Q8LPK5	Q8lpk5 arabidopsis
6	4220	80.5	978	10	Q9AXK0	Q9axk0 zinnia eleg
7	4202	80.2	958	10	Q9SN37	Q9sn37 arabidopsis
8	3699	70.6	939	10	Q943B3	Q943b3 oryza sativ
9	3470	66.2	1043	10	Q9FNC3	Q9fnc3 arabidopsis
10	3464.5	66.1	1055	10	Q8GZN8	Q8gzn8 arabidopsis
11	3406	65.0	1063	10	Q9AV71	Q9av71 oryza sativ
12	3374	64.4	1042	10	O81649	O81649 populus x c
13	3366.5	64.2	1079	10	Q9LLI1	Q9lli1 zea mays (m
14	3366	64.2	1067	10	Q9XGX6	Q9xgx6 gossypium h
15	3365.5	64.2	1077	10	Q9LLI6	Q9lli6 zea mays (m
16	3356.5	64.1	1076	10	Q9LLI5	Q9lli5 zea mays (m
17	3350	63.9	1065	10	Q9FHK6	Q9fhk6 arabidopsis
18	3347	63.9	1081	10	O48946	O48946 arabidopsis
19	3346	63.9	1026	10	Q9SWW6	Q9sww6 arabidopsis
20	3342	63.8	1059	10	Q9LLI4	Q9lli4 zea mays (m
21	3340	63.7	1026	10	Q9XHP6	Q9xhp6 arabidopsis
22	3335.5	63.7	1075	10	Q9LLI9	Q9lli9 zea mays (m
23	3332.5	63.6	1086	10	Q9LLI3	Q9lli3 zea mays (m
24	3332	63.6	1074	10	Q9LLI8	Q9lli8 zea mays (m
25	3332	63.6	1094	10	Q9LLI2	Q9lli2 zea mays (m
26	3327	63.5	1065	10	O48948	O48948 arabidopsis
27	3304	63.1	1032	10	Q8GSW2	Q8gsw2 populus tre
28	3278	62.6	1069	10	Q9FIB9	Q9fib9 arabidopsis
29	3277	62.5	1091	10	Q93XQ1	Q93xq1 nicotiana a
30	3274	62.5	1084	10	O48947	O48947 arabidopsis
31	3251	62.0	1065	10	Q9SKJ5	Q9skj5 arabidopsis
32	3248	62.0	1088	10	Q9SJ22	Q9sj22 arabidopsis
33	3226.5	61.6	1084	10	Q9FGF9	Q9fgf9 arabidopsis
34	3225	61.5	1129	10	Q8LK26	Q8lk26 mesotaenium
35	3199.5	61.1	1081	10	O65338	O65338 arabidopsis
36	3156	60.2	821	10	Q9LLI7	Q9lli7 zea mays (m
37	2730	52.1	685	10	P93156	P93156 gossypium h
38	2249.5	42.9	504	10	Q9AXJ9	Q9axj9 zinnia eleg
39	2234	42.6	1170	10	Q9LHZ7	Q9lhz7 oryza sativ
40	2233.5	42.6	1145	10	Q9LFL0	Q9lfl0 arabidopsis
41	2230	42.6	1104	10	Q8GUZ9	Q8guz9 populus tre
42	2213	42.2	1127	10	Q8W3F9	Q8w3f9 oryza sativ
43	2204.5	42.1	1145	10	Q9M9M4	Q9m9m4 arabidopsis
44	2161	41.2	505	10	Q9AXJ8	Q9axj8 zinnia eleg
45	2157.5	41.2	1111	10	Q9SZL9	Q9szl9 arabidopsis

ALIGNMENTS

RESULT 1
O81368
ID O81368 PRELIMINARY; PRT; 978 AA.
AC O81368;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-JAN-1999 (TrEMBLrel. 09, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Secondary xylem cellulose synthase.
GN CELA.
OS Populus tremuloides (Quaking aspen).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Malpighiales; Salicaceae; Populus.
OX NCBI_TaxID=3693;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Secondary xylem;
RX MEDLINE=20344632; PubMed=10886769;
RA Wu L., Joshi C.P., Chiang V.L.;
RT "A xylem-specific cellulose synthase gene from aspen (Populus
RT tremuloides) is responsive to mechanical stress.";
RL Plant J. 22:495-502(2000).
DR EMBL; AF072131; AAD03417.1; -.
DR InterPro; IPR005150; Cellulose_synt.
DR Pfam; PF03552; Cellulose_synt; 1.
SQ SEQUENCE 978 AA; 110279 MW; C351B0E4975C6B8A CRC64;

Query Match 100.0%; Score 5240; DB 10; Length 978;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 978; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MMESGAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEINEGRKVCLRCGSPYDE 60
              ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db          1 MMESGAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEINEGRKVCLRCGSPYDE 60

Qy         61 NLLDDVEKKGSGNQSTMASHLNDSQDVGIHARHISSVSTVDSEMNDEYGNPIWKNRVKSC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 NLLDDVEKKGSGNQSTMASHLNDSQDVGIHARHISSVSTVDSEMNDEYGNPIWKNRVKSC 120
```

```
Qy        121 KDKENKKKKRSPKAETEPAQVPTEQQMEEKPSAEASEPLSIVYPIPRNKLTPYRAVIIMR 180
Db        121 KDKENKKKKRSPKAETEPAQVPTEQQMEEKPSAEASEPLSIVYPIPRNKLTPYRAVIIMR 180

Qy        181 LVILGLFFHFRITNPVDSAFGLWLTSVICEIWFAFSWVLDQFPKWNPVNRETYIERLSAR 240
Db        181 LVILGLFFHFRITNPVDSAFGLWLTSVICEIWFAFSWVLDQFPKWNPVNRETYIERLSAR 240

Qy        241 YEREGEPSQLAGVDFFVSTVDPLKEPPLITANTVLSILAVDYPVDKVSCYVSDDGAAMLS 300
Db        241 YEREGEPSQLAGVDFFVSTVDPLKEPPLITANTVLSILAVDYPVDKVSCYVSDDGAAMLS 300

Qy        301 FESLVETAEFARKWVPFCKKFSIEPRAPEFYFSQKIDYLKDKVQPSFVKERRAMKRDYEE 360
Db        301 FESLVETAEFARKWVPFCKKFSIEPRAPEFYFSQKIDYLKDKVQPSFVKERRAMKRDYEE 360

Qy        361 YKVRVNALVAKAQKTPEEGWTMQDGTPWPGNNTRDHPGHDSGLPWEILGARDIEGNELPR 420
Db        361 YKVRVNALVAKAQKTPEEGWTMQDGTPWPGNNTRDHPGHDSGLPWEILGARDIEGNELPR 420

Qy        421 LVYVSREKRPGYQHHKKAGAENALVRVSAVLTNAPYILNVDCDHYVNNSKAVREAMCILM 480
Db        421 LVYVSREKRPGYQHHKKAGAENALVRVSAVLTNAPYILNVDCDHYVNNSKAVREAMCILM 480

Qy        481 DPQVGRDVCYVQFPQRFDGIDKSDRYANRNVVFFDVNMKGLDGIQGPVYVGTGCVFNRQA 540
Db        481 DPQVGRDVCYVQFPQRFDGIDKSDRYANRNVVFFDVNMKGLDGIQGPVYVGTGCVFNRQA 540

Qy        541 LYGYGPPSMPSLRKRKDSSSCFSCCCPSKKKPAQDPAEVYRDAKREDLNAAIFNLTEIDN 600
Db        541 LYGYGPPSMPSLRKRKDSSSCFSCCCPSKKKPAQDPAEVYRDAKREDLNAAIFNLTEIDN 600

Qy        601 YDEHERSMLISQLSFEKTFGLSSVFIESTLMENGGVPESANSPPFIKEAIQVIGCGYEEK 660
Db        601 YDEHERSMLISQLSFEKTFGLSSVFIESTLMENGGVPESANSPPFIKEAIQVIGCGYEEK 660

Qy        661 TEWGKQIGWIYGSVTEDILSGFKMHCRGWRSIYCMPVRPAFKGSAPINLSDRLHQVLRWA 720
Db        661 TEWGKQIGWIYGSVTEDILSGFKMHCRGWRSIYCMPVRPAFKGSAPINLSDRLHQVLRWA 720

Qy        721 LGSVEIFFSRHCPLWYGFGGGRLKWLQRLAYINTIVYPFTSLPLIAYCTIPAVCLLTGKF 780
Db        721 LGSVEIFFSRHCPLWYGFGGGRLKWLQRLAYINTIVYPFTSLPLIAYCTIPAVCLLTGKF 780

Qy        781 IIPTLSNLASMLFLGLFISIIVTAVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQG 840
Db        781 IIPTLSNLASMLFLGLFISIIVTAVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQG 840

Qy        841 FLKMLAGIDTNFTVTAKAAEDAEFGELYMVKWTTLLIPPTTLLIINMSGCAGFSDALNKG 900
Db        841 FLKMLAGIDTNFTVTAKAAEDAEFGELYMVKWTTLLIPPTTLLIINMSGCAGFSDALNKG 900

Qy        901 YEAWGPLFGKVFFAFWVILHLYPFLKGLMGRQNLTPTIVVLWSVLLASVFSLVWVKINPF 960
Db        901 YEAWGPLFGKVFFAFWVILHLYPFLKGLMGRQNLTPTIVVLWSVLLASVFSLVWVKINPF 960

Qy        961 VNKVDNTLVAETCISIDC 978
Db        961 VNKVDNTLVAETCISIDC 978
```

```
ESULT 2
93155
D   P93155      PRELIMINARY;      PRT;   974 AA.
C   P93155;
T   01-MAY-1997 (TrEMBLrel. 03, Created)
T   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
T   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
E   Cellulose synthase.
N   CELA1.
S   Gossypium hirsutum (Upland cotton).
C   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
C   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium.
OX   NCBI_TaxID=3635;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Acala SJ 2; TISSUE=Fiber;
RX   MEDLINE=97057296; PubMed=8901636;
RA   Pear J.R., Kawagoe Y., Schreckengost W.E., Delmer D.P., Stalker D.M.;
RT   "Higher plants contain homologs of the bacterial celA genes encoding
RT   the catalytic subunit of cellulose synthase.";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:12637-12642(1996).
DR   EMBL; U58283; AAB37766.1;
DR   InterPro; IPR005150; Cellulose_synt.
DR   Pfam; PF03552; Cellulose_synt; 1.
SQ   SEQUENCE   974 AA;  109703 MW;  2656CA2D35D109A7 CRC64;

  Query Match             87.3%;  Score 4524;  DB 10;  Length 974;
  Best Local Similarity   87.5%;  Pred. No. 0;
  Matches  840;  Conservative   70;  Mismatches   60;  Indels   12;  Gaps    7;

Qy          1 MMESGAFTHTT  AGRDANGTPVACHEGYPMCKSCFEFEINEGRKVCLRCGSPYDE 60
Db          1 MMESGVFTHTT  PVTDVNGTPVACHEGYPICKSCFEYDLKEGRKACLRCGSPYDE 60

Qy         61 NLLDDVEKKGSGNQSTMAASHLNKSQDVGIHARHISSVSTVDSEMNDEYGNPIWKNRVKSC 120
Db         61 NLLDDVE-KATGDQSTMAAHLNKSQDVGIHARHISSVSTLDSEMAEDNGNSIWKNRVESW 119

Qy        121 KDKENKKKKRSPEAETEPAQVPTEQQMEEKPSAEASEPLSIVYPIPRNKLTPYRAVIIMR 180
Db        120 KEKKNKKKKPATTPVEREAEIPTEQQMEDKPAPDASQPLSTIIPIPKSRLAPYRTVIIMR 179

Qy        181 LVILGLFFHFRITNPVDSAFGLWLTSVICEIWFAFSWVLDQFPKWNPVNRETYIERLSAR 240
Db        180 LIILGLFFHYRVTNPVDSAFGLWLTSVICEIWFAFSWVLDQFPKWYPVNRETYIDRLSAR 239

Qy        241 YEREGEPSQLAGVDFFVSTVDPLKEPPLITANTVLSILAVDYPVDKVSCYVSDDGAAMLS 300
Db        240 YEREGEPSQLAAVDFFVSTVDPLKEPPLITANTVLSILALDYPVDKVSCYISDDGAAMLT 299

Qy        301 FESLVETAEFARKWVPFCKKFSIEPRAPEFYFSQKIDYLKDKVQPSFVKERRAMKRDYEE 360
Db        300 FESLVETAEFARKWVPFCKKFSIEPRAPEFYFSQKIDYLKDKVQPSFVKERRAMKRDYEE 359

Qy        361 YKVRVNALVAKAQKTPEEGWTMQDGTPWPGNNTRDHPGHDSGLPWEIL---GARDIEGNE 417
Db        360 YKIRINALVAKAQKTPDEGWTMQDGTSWPGNNPRDHP----GMIQVFLGYSGARDIEGNE 415

Qy        418 LPRLVYVSREKRPGYQHHKKAGAENALVRVSAVLTNAPYILNVDCDHYVNNSKAVREAMC 477
Db        416 LPRLVYVSPEKRPGYQHHKKAGAENALVRVSAVLTNAPFILNLDCDHYVNNSKAVREAMC 475

Qy        478 ILMDPQVGRDVCYVQFPQRFDGIDKSDRYANRNVVFFDVNMKGLDGIQGPVYVGTGCVFN 537
Db        476 FLMDPQVGRDVCYVQFPQRFDGIDRSDRYANRNTVFFDVNMKGLDGIQGPVYVGTGCVFN 535

Qy        538 RQALYGYGPPSMPSLRKRKDSSSCFSCCCPSKKKPAQDPAEVYRDAKREDLNAAIFNLTE 597
Db        536 RQALYGYGPPSMPSFPK-SSSSSC-SCCCPGKKEP KDPSELYRDAKREELDAAIFNLRE 592

Qy        598 IDNYDEHERSMLISQLSFEKTFGLSSVFIESTLMENGGVPESANSPPFIKEAIQVIGCGY 657
Db        593 IDNYDEYERSMLISQTSFEKTFGLSSVFIESTLMENGGVAESANPSTLIKEAIHVISCGY 652

Qy        658 EEKTEWGKQIGWIYGSVTEDILSGFKMHCRGWRSIYCMPVRPAFKGSAPINLSDRLHQVL 717
Db        653 EEKTAWGKEIGWIYGSVTEDILTGFKMHCRGWRSIYCMPLRPAFKGSAPINLSDRLHQVL 712

Qy        718 RWALGSVEIFFSRHCPLWYGFGGGRLKWLQRLAYINTIVYPFTSLPLIAYCTIPAVCLLT 777
Db        713 RWALGSVEIFLSRHCPLWYGFGGGRLKWLQRLAYINTIVYPFTSLPLIAYCSLPAICLLT 772

Qy        778 GKFIIPTLSNLASMLFLGLFISIIVTAVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAV 837
```

```
Db        773 GKFIIPTLSNLASVLFLGLFLSIIVTAVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAV 832

Qy        838 FQGFLKMLAGIDTNFTVTAKAAEDAEFGELYMVKWTTLLIPPTTLLIINMSG-CAGFSDA 896
              ||||||||||||||||||||||:||:|||||:|||||||||||||||||| | ||||||
Db        833 FQGFLKMLAGIDTNFTVTAKAADDADFGELYIVKWTTLLIPPTTLLIVNMVGVVAGFSDA 892

Qy        897 LNKGYEAWGPLFGKVFFAFWVILHLYPFLKGLMGRQNLTPTIVVLWSVLLASVFSLVWVK 956
              ||||||||||||||||||:|||||||||||||||:||||| ||||||||||:|||||||||:
Db        893 LNKGYEAWGPLFGKVFFSFWVILHLYPFLKGLMGRQNRTPTIVVLWSVLLASVFSLVWVR 952

Qy        957 INPFVNKVDNTLVAETCISIDC 978
              |||||:  |:| |:::||||||
Db        953 INPFVSTADSTTVSQSCISIDC 974


RESULT 3
Q8W1W0
ID   Q8W1W0      PRELIMINARY;      PRT;   974 AA.
AC   Q8W1W0;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Cellulose synthase A4.
GN   CESA4.
OS   Gossypium hirsutum (Upland cotton).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium.
OX   NCBI_TaxID=3635;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21608363; PubMed=11743074;
RA   Kim H.J., Triplett B.A.;
RT   "Cotton Fiber Growth in Planta and in Vitro. Models for Plant Cell
RT   Elongation and Cell Wall Biogenesis.";
RL   Plant Physiol. 127:1361-1366(2001).
DR   EMBL; AF413210; AAL37718.1; -.
DR   InterPro; IPR005150; Cellulose_synt.
DR   Pfam; PF03552; Cellulose_synt; 1.
SQ   SEQUENCE   974 AA;  109499 MW;  F8AFB66ABA2B8071 CRC64;

  Query Match             84.5%;  Score 4426;  DB 10;  Length 974;
  Best Local Similarity   83.8%;  Pred. No. 0;
  Matches  823;  Conservative   75;  Mismatches   72;  Indels   12;  Gaps    7;

Qy          1 MMESGAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEINEGRKVCLRCGSPYDE 60
              |||| |:|||||| || : '|| |.'||||::|.|: ||| : |||' ||||||||||
Db          1 MMESGVPVCHTCGEHVGLNVNGEPFVACHECNFPICKSCFEYELKEGRKACLRCGSPYDE 60

Qy         61 NLLDDVEKKGSGNQSTMAAHLNDSIQVGIHARHISSVSTVL-MNDEYGNPIWKNRVKSC 120
              |||||||  :|:|||||:|| : ||||:||| ||||: || : ||||||||:|
Db         61 NLLDDVEKT-TGDQSTMAAHLSKSQDVGIHARHISSVSTLD-EMTGDNGNPIWKNRVESW 119

Qy        121 KDKENKKKKRSPKAETEPAQVPTEQQMEEKPSAEASEPLSIVYPIPRNKLTPYRAVIIMR 180
              |:|:|||||  :       |::| ||||:||: :||:|||   |||:::| ||| |||||
Db        120 KEKKNKKKKKPATTKVEREAEIPPEQQMEDKPAPDASQPLST;IPIPKSRLAPYRTVIIMR 179

Qy        181 LVILGLFFHFRITNPVDSAFGLWLTSVICEIWFAFSWVLDQ:PKWNPVNRETYIERLSAR 240
              |:|||||||:|:|:||||||||||||||||||||:|||:|||:|| |||'||||:|||||
Db        180 LIILGLFFHYRVTHPVDSAFGLWLTSVICEIWFAFSWVLDQ:PKWYPVNRETYIDRLSAR 239

Qy        241 YEREGEPSQLAGVDFFVSTVDPLKEPPLITANTVLSILAVD:PVDKVSCYVSDDGAAMLS 300
              |:||||::|| |||||||||||||||||||||||:|:|:| :|||||||||||||||||:
Db        240 YDREGEPNELAAVDFFVSTVDPLKEPPLITANTVLSILALDYPVDKVSCYISDDGAAMLT 299

Qy        301 FESLVETAEFARKWVPFCKKFSIEPRAPEFYFSQKIDYLKDKVQPSFVKERRAMKRDYEE 360
              ||||||||:|||||||||||||||||||||:||||| |||| |||||||||||||||||
Db        300 FESLVETADFARKWVPFCKKFSIEPRAPEFYFSQKIVYLKDRVQPSFVKERRAMKRDYEE 359

Qy        361 YKVRVNALVAKAQKTPEEGWTMQDGTPWPGNNTRDHPGHDSGLPWEIL---GARDIEGNE 417
              ||:|:|||||||||||||||.|||'|||| | ||    }:    |   || ||||||
Db        360 YKIRINALVAKAQKTPEEGWTMQDGTPWPGNNPRDHP----GMIQVFLGYSGAHDIEGNE 415
```

```
Qy        418 LPRLVYVSREKRPGYQHHKKAGAENALVRVSAVLTNAPYILNVDCDHYVNNSKAVREAMC 477
              |||||||||||||||||||:|||:|||||||| |||||:|||:||||||||||||||||
Db        416 LPRLVYVSREKRPGYQHHKQAGADNALVRVSAALTNAPFILNLDCDHYVNNSKAVREAMC 475

Qy        478 ILMDPQVGRDVCYVQFPQRFDGIDKSDRYANRNVVFFDVNMKGLDGIQGPVYVGTGCVFN 537
              |||||:|||||||||||||:|||:|:||| || |||||||||| || ||||||||||| |
Db        476 CLMDPQGGRDVCYVQFPQRCDGIDRSERYAKRNTVFFDVNMKGRDGSQGPVYVGTGCVCN 535

Qy        538 RQALYGYGPPSMPSLRKRKDSSSCFSCCCPSKKKPAQDPAEVYRDAKREDLNAAIFNLTE 597
              |||||||||||||:|  |:|||| ||||| ||:| ::| |:|||||||:|:|||||| |
Db        536 RQALYGYGPPSMPSFPK-SSSSSC-SCCCPGKKEP-KEPTELYRDAKREELDAAIFNLRE 592

Qy        598 IDNYDEHERSMLISQLSFEKTFGLSSVFIESTLMENGGVPESANSPPFIKEAIQVIGCGY 657
              ||||||:|||||||| ||| ||||||||||||||||| ||||    ||||| ||||||
Db        593 IDNYDEYERSMLISQTSFEKTFGLSSVFIESTLMENGGVAESANPSTLIKEAIHVIGCGY 652

Qy        658 EEKTEWGKQIGWIYGSVTEDILSGFKMHCRGWRSIYCMPVRPAFKGSAPINLSDRLHQVL 717
              |||| |||:|||||||||||||:|||||||||||||||||:|||||||||||||||||||
Db        653 EEKTAWGKEIGWIYGSVTEDILTGFKMHCRGWRSIYCMPLRPAFKGSAPINLSDRLHQVL 712

Qy        718 RWALGSVEIFFSRHCPLWYGFGGGRLKWLQRLAYINTIVYPFTSLPLIAYCTIPAVCLLT 777
              ||||||:||| |||||||||||||||||||||||||| ||||||||||||::||:||||
Db        713 RWALGSLEIFLSRHCPLWYGFGGGRLKWLQRLAYINTSVYPFTSLPLIAYCSLPAICLLT 772

Qy        778 GKFIIPTLSNLASMLFLGLFISIIVTAVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAV 837
              |||||||||||||:|:||||:||||||||||||||||||||||||||||||||||||||||
Db        773 GKFIIPTLSNLASVLYLGLFLSIIVTAVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAV 832

Qy        838 FQGFLKMLAGIDTNFTVTAKAAEDAEFGELYMVKWTTLLIPPTTLLIINMSG-CAGFSDA 896
              |||||||||||||||||||:||:|||||:|||||||||||||||:|| | ||||||
Db        833 FQGFLKMLAGIDTNFTVTAKAADDADFGELYIVKWTTLLIPPTTLLIVNMVGVVAGFSDA 892

Qy        897 LNKGYEAWGPLFGKVFFAFWVILHLYPFLKGLMGRQNLTPTIVVLWSVLLASVFSLVWVK 956
              ||||||||||||||||||:|||||||||||||||:||||| ||||||||||:|||||||||:
Db        893 LNKGYEAWGPLFGKVFFSFWVILHLYPFLKGLMGRQNRTPTIVVLWSVLLASVFSLVWVR 952

Qy        957 INPFVNKVDNTLVAETCISIDC 978
              |||||:  |:| |:    ||||
Db        953 INPFVSTADSTTVSRANSSIDC 974


RESULT 4
Q9C5Z8
ID   Q9C5Z8      PRELIMINARY;      PRT;   985 AA.
AC   Q9C5Z8;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Cellulose synthase catalytic subunit.
GN   IRX1.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21066939; PubMed=11148295;
RA   Taylor N.G., Laurie S., Turner S.R.;
RT   "Multiple Cellulose Synthase Catalytic Subunits Are Required for
RT   Cellulose Synthesis in Arabidopsis.";
RL   Plant Cell 12:2529-2540(2000).
DR   EMBL; AF267742; AAK08700.1; -.
DR   InterPro; IPR005150; Cellulose_synt.
DR   Pfam; PF03552; Cellulose_synt; 1.
SQ   SEQUENCE   985 AA;  111470 MW;  693C701B7B69AB25 CRC64;

  Query Match             82.5%;  Score 4323.5;  DB 10;  Length 985;
  Best Local Similarity   81.0%;  Pred. No. 0;
  Matches  804;  Conservative   89;  Mismatches   78;  Indels   21;  Gaps    8;

Qy          1 MMESGAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEINEGRKVCLRCGSPYDE 60
              |||| :|||:||||::|  :||| ||||||||:|:||:| |:|  |||::|||||:||||
```

```
  1 MMESRSPICNTCGEEIGAKSNGEFFVACHECSFPICKACLEYEFKEGRRICLRCGNPYDE 60

 61 NLLDDVEKKGSGNQSTMASHLND-SQDVGIHARHISSVSTVDSEMNDEYGNPIWKNRVKS 119
    |: |||| | |  || : :  |: ||| |||||||:|||:|||:|||||||||||||:| 
 61 NVFDDVETKTSKTQSIVPTQTNNTSQDSGIHARHISTVSTIDSELNDEYGNPIWKNRVES 120

120 C---KDKENKKKKRSP---KAETEPAQVPTEQQMEEKP---SAEASEPLSIVYPIPRNKL 170
       |||::||||: |   ||| : |||||:| ||: |    : |:: ||:| |||| |: 
121 WKDKKDKKSKKKKKDPKATKAEQQDAQVPTQQHMEDMPPNTESGATDVLSVVIPIPRTKI 180

171 TPYRAVIIMRLVILGLFFHFRITNPVDSAFGLWLTSVICEIWFAFSWVLDQFPKWNPVNR 230
    | || ||||||:|| |||::|||:|||||:|||||||||||||| ||||||||||:|:|| 
181 TSYRIVIIMRLIILALFFNYRITHPVDSAYGLWLTSVICEIWFAVSWVLDQFPKWSPINR 240

231 ETYIERLSARYEREGEPSQLAGVDFFVSTVDPLKEPPLITANTVLSILAVDYPVDKVSCY 290
    ||||:|||||:||||| |||| |||||||||||||||||||||||||||:|||||||||| 
241 ETYIDRLSARFEREGEQSQLAAVDFFVSTVDPLKEPPLITANTVLSILALDYPVDKVSCY 300

291 VSDDGAAMLSFESLVETAEFARKWVPFCKKFSIEPRAPEFYFSQKIDYLKDKVQPSFVKE 350
    |||||||||||||||:|||||||:|||:|:|||:|||||| ||||:|||||||||||||| 
301 VSDDGAAMLSFESLVETADFARKWVPFCKKYSIEPRAPEFYFSLKIDYLRDKVQPSFVKE 360

351 RRAMKRDYEEYKVRVNALVAKAQKTPEEGWTMQDGTPWPGNNTRDHPGHDSGLPWEIL-- 408
    ||||||||||:|:|:|||||||||||||||||||| |||||||||     |:    | 
361 RRAMKRDYEEFKIRMNALVAKAQKTPEEGWTMQDGTSWPGNNTRDHP----GMIQVFLGY 416

409 -GARDIEGNELPRLVYVSREKRPGYQHHKKAGAENALVRVSAVLTNAPYILNVDCDHYVN 467
     |||||||||||||||||||||||||||||||||||||||||||||||:|||:|||||||| 
417 SGARDIEGNELPRLVYVSREKRPGYQHHKKAGAENALVRVSAVLTNAPFILNLDCDHYVN 476

468 NSKAVREAMCILMDPQVGRDVCYVQFPQRFDGIDKSDRYANRNVVFFDVNMKGLDGIQGP 527
    |||||||||| |||:|||||||||||||||||||||||||:||||||:|||:|||||||| 
477 NSKAVREAMCFLMDPVVGQDVCFVQFPQRFDGIDKSDRYANRNIVFFDVNMRGLDGIQGP 536

528 VYVGTGCVFNRQALYGYGPPSMPSLRKRKDSSSCFSCCCPSKKKPAQDPAEVYRDAKRED 587
    |||||| || ||||||| ||| | :  :  ||   |||| :|||  |||||:|:|||||: 
537 VYVGTGTVFRRQALYGYSPPSKPRILPQSSSS---SCCCLTKKKQPQDPAEIYKDAKREE 593

588 LNAAIFNLTEIDNYDEHERSMLISQLSFEKTFGLSSVFIESTLMENGGVPESANSPPFIK 647
    |:|||||| ::|||||::|||||| |||||||||:|||||||||||||:| |    || 
594 LDAAIFNLGDLDNYDEYDRSMLISQTSFEKTFGLSTVFIESTLMENGGVPDSVNPSTLIK 653

648 EAIQVIGCGYEEKTEWGKQIGWIYGSVTEDILSGFKMHCRGWRSIYCMPVRPAFKGSAPI 707
    ||| || ||||||||||:|||||||:|||||:|||||||||||||||||:|||||||||| 
654 EAIHVISCGYEEKTEWGKEIGWIYGSITEDILTGFKMHCRGWRSIYCMPLRPAFKGSAPI 713

708 NLSDRLHQVLRWALGSVEIFFSRHCPLWYGFGGGRLKWLQRLAYINTIVYPFTSLPLIAY 767
    |||||||||||||||||| |||||||||  ||||| |||||||||||||||||||:||| 
714 NLSDRLHQVLRWALGSVEIFLSRHCPLWYGCSGGRLKLLQRLAYINTIVYPFTSLPLVAY 773

768 CTIPAVCLLTGKFIIPTLSNLASMLFLGLFISIIVTAVLELRWSGVSIEDLWRNEQFWVI 827
    ||:||:|||||||||||||||||||||||||:|:|||||||||||||||||||||||||| 
774 CTLPAICLLTGKFIIPTLSNLASMLFLGLFISIILTSVLELRWSGVSIEDLWRNEQFWVI 833

828 GGVSAHLFAVFQGFLKMLAGIDTNFTVTAKAAEDAEFGELYMVKWTTLLIPPTTLLIINM 887
    |||||||||||||||||||:|||||||:| |:| |||||:|||||||||||:|||||: 
834 GGVSAHLFAVFQGFLKMLAGLDTNFTVTSKTADDLEFGELYIVKWTTLLIPPTSLLIINL 893

888 SG-CAGFSDALNKGYEAWGPLFGKVFFAFWVILHLYPFLKGLMGRQNLTPTIVVLWSVLL 946
      | |||||||||||||||||||||||||||||||||||||||||||||| |||||:|||:|| 
894 VGVVAGFSDALNKGYEAWGPLFGKVFFAFWVILHLYPFLKGLMGRQNRTPTIVILWSILL 953

947 ASVFSLVWVKINPFVNKVDNTLVAETCISIDC 978
    |||||||||:|||||:| | | ::  |: ||| 
954 ASVFSLVWVRINPFVSKTDTTSLSLNCLLIDC 985
```

```
ESULT 5
Q8LPK5
ID   Q8LPK5      PRELIMINARY;      PRT;   985 AA.
AC   Q8LPK5;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Cellulose synthase-like protein.
GN   AT4G18780.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Nguyen M., Karlin-Neumann G., Southwick A., Lam B., Miranda M.,
RA   Palm C.J., Bowser L., Jones T., Bahk J., Carninci P., Chen H.,
RA   Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA   Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA   Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
RA   Ecker J., Theologis A., Davis R.W.;
RL   Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY099636; AAM20748.1; -.
DR   InterPro; IPR003330; de_lucose_synt.
DR   Pfam; PF03552; Cellulose_synt; 1.
SQ   SEQUENCE   985 AA;  [illegible] MW;  12[illegible]5768E3 CRC64;

  Query Match             8.4%;  Score 4316.5;  DB 10;  Length 985;
  Best Local Similarity   80.8%;  Pred. No. 0;
  Matches  802;  Conservative   90;  Mismatches   79;  Indels   21;  Gaps    8;

Qy          1 MMESGAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEINEGRKVCLRCGSPYDE 60
              |||| :|:::|||::|  :||| ||||||.|:|:|':| |:|  |||::|||||:|||
Db          1 MMESRSPICNTCGEEIGVKSNGEFFVACHECSFPICKACLEYEFKEGRRICLRCGNPYDE 60

Qy         61 NLLDDVEKKGSGNQSTMASHLND-SQDVGIHARHISSVSTVDSEMNDEYGNPIWKNRVKS 119
              |: |||| | |  || : :  |: ||| |||||||:|||:|||:|||||||||||||:| 
Db         61 NVFDDVETKTSKTQSIVPTQTNNTSQDSGIHARHISTVSTIDSELNDEYGNPIWKNRVES 120

Qy        120 C---KDKENKKKKRSP---KAETEPAQVPTEQQMEEKP---SAEASEPLSIVYPIPRNKL 170
                 |||::||||: |   ||| : |||||:| ||: |    : |:: ||:| |||| |: 
Db        121 WKDKKDKKSKKKKKDPKATKAEQQDAQVPTQQHMEDMPPNTESGATDVLSVVIPIPRTKI 180

Qy        171 TPYRAVIIMRLVILGLFFHFRITNPVDSAFGLWLTSVICEIWFAFSWVLDQFPKWNPVNR 230
              | || ||||||:|| |||::|||:|||||:|||||||||||||| ||||||||||:|:|| 
Db        181 TSYRIVIIMRLIILALFFNYRITHPVDSAYGLWLTSVICEIWFAVSWVLDQFPKWSPINR 240

Qy        231 ETYIERLSARYEREGEPSQLAGVDFFVSTVDPLKEPPLITANTVLSILAVDYPVDKVSCY 290
              ||||:|||||:||||| |||| |||||||||||||||||||||||||||:|||||||||| 
Db        241 ETYIDRLSARFEREGEQSQLAAVDFFVSTVDPLKEPPLITANTVLSILALDYPVDKVSCY 300

Qy        291 VSDDGAAMLSFESLVETAEFARKWVPFCKKFSIEPRAPEFYFSQKIDYLKDKVQPSFVKE 350
              |||||||||||||||:|||||||:|||:|:|||:|||||| ||||:|||||||||||||| 
Db        301 VSDDGAAMLSFESLVETADFARKWVPFCKKYSIEPRAPEFYFSLKIDYLRDKVQPSFVKE 360

Qy        351 RRAMKRDYEEYKVRVNALVAKAQKTPEEGWTMQDGTPWPGNNTRDHPGHDSGLPWEIL-- 408
              ||||||||||:|:|:|||||||||||||||||||| |||||||||     |:    | 
Db        361 RRAMKRDYEEFKIRMNALVAKAQKTPEEGWTMQDGTSWPGNNTRDHP----GMIQVFLGY 416

Qy        409 -GARDIEGNELPRLVYVSREKRPGYQHHKKAGAENALVRVSAVLTNAPYILNVDCDHYVN 467
               |||||||||||||||||||||||||||||||||||||||||||||||:|||:|||||||| 
Db        417 SGARDIEGNELPRLVYVSREKRPGYQHHKKAGAENALVRVSAVLTNAPFILNLDCDHYVN 476

Qy        468 NSKAVREAMCILMDPQVGRDVCYVQFPQRFDGIDKSDRYANRNVVFFDVNMKGLDGIQGP 527
              |||||||||| |||:|||||||||||||||||||||||||:||||||:|||:|||||||| 
Db        477 NSKAVREAMCFLMDPVVGQDVCFVQFPQRFDGIDKSDRYANRNIVFFDVNMRGLDGIQGP 536

Qy        528 VYVGTGCVFNRQALYGYGPPSMPSLRKRKDSSSCFSCCCPSKKKPAQDPAEVYRDAKRED 587
              |||||| || ||||||| ||| | :  :  ||   |||| :|||  |||||:|:|||||: 
Db        537 VYVGTGTVFRRQALYGYSPPSKPRILPQSSSS---SCCCLTKKKQPQDPAEIYKDAKREE 593

Qy        588 LNAAIFNLTEIDNYDEHERSMLISQLSFEKTFGLSSVFIESTLMENGGVPESANSPPFIK 647
              |:|||||| ::|||:|::|||||| |||:|||||:|||||||||||||:| |    || 
Db        594 LDAAIFNLGDLDNYDEYDRSMLISQTSFEKTFGLSTVFIESTLMENGGVPDSVNPSTLIK 653

Qy        648 EAIQVIGCGYEEKTEWGKQIGWIYGSVTEDILSGFKMHCRGWRSIYCMPVRPAFKGSAPI 707
              ||| || ||||||||||:|||||||:|||||:|||||||||||||||||:|||||||||| 
Db        654 EAIHVISCGYEEKTEWGKEIGWIYGSITEDILTGFKMHCRGWRSIYCMPLRPAFKGSAPI 713
```

```
Qy        708 NLSDRLHQVLRWALGSVEIFFSRHCPLWYGFGGGRLKWLQRLAYINTIVYPFTSLPLIAY 767
              ||||||||||||||||||| |||||||| ||||| |||:||||||||||||||||:||
Db        714 NLSDRLHQVLRWALGSVEIFLSRHCPLWYGCSGGRLKLLQRLAYINTIVYPFTSLPLVAY 773

Qy        768 CTIPAVCLLTGKFIIPTLSNLASMLFLGLFISIIVTAVLELRWSGVSIEDLWRNEQFWVI 827
              ||:||:|||||||||||||||||:|||||||:|:|||||||||||||||||||||||||
Db        774 CTLPAICLLTGKFIIPTLSNLASMLFLGLFISIILTSVLELRWSGVSIEDLWRNEQFWVI 833

Qy        828 GGVSAHLFAVFQGFLKMLAGIDTNFTVTAKAAEDAEFGELYMVKWTTLLIPPTTLLIINM 887
              |||||||||||||||||||:|:|||||:| |:| ||||| ||||||||||:|||||:
Db        834 GGVSAHLFAVFQGFLKMLAGLDTNFTVTSKTADDLEFGELYIVKWTTLLIPPTSLLIINL 893

Qy        888 SG-CAGFSDALNKGYEAWGPLFGKVFFAFWVILHLYPFLKGLMGRQNLTPTIVVLWSVLL 946
              |  |||||||||||||||||||||||||||||||||||||||||| ||||||:|||:||
Db        894 VGVVAGFSDALNKGYEAWGPLFGKVFFAFWVILHLYPFLKGLMGRQNRTPTIVILWSILL 953

Qy        947 ASVFSLVWVKINPFVNKVDNTLVAETCISIDC 978
              |||||||||:|||||:| | | ::  |: |||
Db        954 ASVFSLVWVRINPFVSKTDTTSILSLNCLLIDC 985


RESULT 6
Q9AXK0
ID   Q9AXK0      PRELIMINARY;      PRT;   978 AA.
AC   Q9AXK0;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Cellulose synthase CesA-1.
OS   Zinnia elegans.
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC   Asteridae; campanulids; Asterales; Asteraceae; Asteroideae;
OC   Heliantheae; Zinnia.
OX   NCBI TaxID=34245;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21438255; PubMed=11554477;
RA   Haigler C.H., Ivanova-Datcheva M., Hogan P.S., Salnikov V.V.,
RA   Hwang S., Martin K., Delmer D.P.;
RT   "Carbon partitioning to cellulose synthesis.";
RL   Plant Mol. Biol. 47:29-51(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Hwang S., Haigler C.H.;
RL   Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF323039; AAK11588.2; -.
DR   InterPro; IPR005150; Cellulose_synt.
DR   Pfam; PF03552; Cellulose_synt; 1.
SQ   SEQUENCE   978 AA;  110053 MW;  399D76EF3342/BID CRC64;

  Query Match           80.5%;  Score 4220;  DB 10;  Length 978;
  Best Local Similarity   79.4%;  Pred. No. 0;
  Matches  788;  Conservative   90;  Mismatches   85;  Indels   30;  Gaps   12;

Qy          1 MMESGAPICHTCGEQVGH--DANGELFVACHECSYPMCKSCFEFEINEGRKVCLRCGSPY 58
              ||    |||:|||| :|   | || :|| |:||::|:|||| |:|||||   |::||:||
Db          1 MMHPETPICNTCGEHIGFGPDGNG-VFVGCHECNFPICKSCIEYEINEGHTACIQCGTPY 59

Qy         59 DENL--LDDVEKKGSGNQSTMASHLNDSQDVGIHARHISSVSTVDSEMNDEYGNPIWKNR 116
              | ||  : ::||:    ::|||:|::||| |:|||::|; |||:|| || ||||||||
Db         60 DGNLTKVGEMEKE-PVTRNTMAAHLSNSQDAGLHARNVSTVSTVDTEMTDESGNPIWKNR 118

Qy        117 VKSCKDKENKKKKRSPKAETEPAQVPTEQQMEEKPSA---EASEPLSIVYPIPRNKLTPY 173
              |:| |||:|||||   |   |  |:| :| :|||  :   : :| | : |||::::|||
Db        119 VESWKDKKNKKKKGGDKVAKE-VQVPEDQHIEEKQCSADPNAMQPPSQIIPIPKSQITPY 177

Qy        174 RAVIIMRLVILGLFFHFRITNPVDSAFGLWLTSVICEIWFAFSWVLDQFPKWNPVNRETY 233
              | |||:||:|||||||:||||||:|::||:||||||||| ||||||||| |:|| |:
Db        178 RIVIILRLIILGLFFHYRITNPVESSYGLWLTSVICEIWFAISWVLDQFPKWYPINRVTF 237

Qy        234 IERLSARYEREGEPSQLAGVDFFVSTVDPLKEPPLITANTVLSILAVDYPVDKVSCYVSD 293
              : ||||||||||:|| ||||||||||||||||||||||||||||:||||||||
Db        238 TDELSARYEREGEPSELAAVDFFVSTVDPLKEPPLITANTVLSILAVDYPVEKVSCYVSD 297

Qy        294 DGAAMLSFESLVETAEFARKWVPFCKKFSIEPRAPEFYFSQKIDYLKDKVQPSFVKERRA 353
              |||||:|||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        298 DGAAMLTFESLVETAEFARKWVPFCKKFSIEPRAPEFYFSQKIDYLKDKVQPSFVKERRA 357

Qy        354 MKRDYEEYKVRVNALVAKAQKTPEEGWTMQDGTPWPGNNTRDHP-------GHDSGLPWE 406
              |||||||||||||||||| |||||||||||||||||| ||||        ||
Db        358 MKRDYEEYKVRVNALVAKAQKTPEEGWTMQDGTPWPGNNPRDHPGMIQVFLGHS------ 411

Qy        407 ILGARDIEGNELPRLVYVSREKRPGYQHHKKAGAENALVRVSAVLTNAPYILNVDCDHYV 466
              || |||||||||||||:|||||||||||||||||||||||||||||||||||:||||||
Db        412 --GAHDIEGNELPRLVYVSREKRPGYQHHKKAGAENALVRVSAVLTNAPYILNLDCDHYV 469

Qy        467 NNSKAVREAMCILMDPQVGRDVCYVQFPQRFDGIDKSDRYANRNVVFFDVNMKGLDGIQG 526
              |||||||||| :|||||||||:|||||||||:|||||||| ||||||||||||||
Db        470 NNSKAVREAMCFMMDPQVGRDVCYIQFPQRFDGIDRSDRYANRNTVFFDVNMKGLDGIQG 529

Qy        527 PVYVGTGCVFNRQALYGYGPPSMPSLRKRKDSSSCFSCCCPSKKKPAQDPAEVYRDAKRE 586
              |||||||||| |||||||| |:|:|     |||| |||   ||| :|  |  |||:|:
Db        530 PVYVGTGCVFYRQALYGYGPQSLPTLPSPSSSSSC--CCC-GPKKPKKDLEEFKRDARRD 586

Qy        587 DLNAAIFNLTEIDNYDEHERSMLISQLSFEKTFGLSSVFIESTLMENGGVPESANSPPFI 646
              |||||||||  ||::||::|||:||||:|||||||:|||||||||||||: ||||    |
Db        587 DLNAAIFNLKEIESYDDYERSLLISQMSFEKTFGMSSVFIESTLMENGGLAESANPATMI 646

Qy        647 KEAIQVIGCGYEEKTEWGKQIGWIYGSVTEDILSGFKMHCRGWRSIYCMPVRPAFKGSAP 706
              |||  ||  |||||||  |||:|||||||||||||:|||||||||||||||||||||||
Db        647 NEAIHVISCGYEEKTAWGKEIGWIYGSVTEDILTGFKMHCRGWRSIYCMPVRPAFKGSAP 706

Qy        707 INLSDRLHQVLRWALGSVEIFFSRHCPLWYGFGGGRLKWLQRLAYINTIVYPFTSLPLIA 766
              |||||||||||||||||| |||||||||:|||||| ||||||||||||||||||:|
Db        707 INLSDRLHQVLRWALGSVEIFLSRHCPLWYGWGGGRLKLLQRLAYINTIVYPFTSLPLVA 766

Qy        767 YCTIPAVCLLTGKFIIPTLSNLASMLFLGLFISIIVTAVLELRWSGVSIEDLWRNEQFWV 826
              |||:||:|||||||||||||:|:: |||||:||| |:|||:|||||||:|||||||||||
Db        767 YCTLPAICLLTGKFIIPTLSNIAAVWFLGLFLSIITTSVLEIRWSGVSIEELWRNEQFWV 826

Qy        827 IGGVSAHLFAVFQGFLKMLAGIDTNFTVTAKAAEDAEFGELYMVKWTTLLIPPTTLLIIN 886
              |||||||||||||||||||:|||||||||||:| |||||||:||||:|||||||:|||||
Db        827 IGGVSAHLFAVFQGFLKMLAGVDTNFTVTAKAADDQEFGELYMIKWTTVLIPPTTLLVLN 886

Qy        887 MSG-CAGFSDALNKGYEAWGPLFGKVFFAFWVILHLYPFLKGLMGRQNLTPTIVVLWSVL 945
              : |  |||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db        887 LVGVVAGFSDALNKGYEAWGPLFGKVFFAFWVILHLYPFLKGLMGRQNRTPTIVILWSVL 946

Qy        946 LASVFSLVWVKINPFVNKVDNTLVAETCISIDC 978
              |||||||||||:|||:| |: |  :  ||:|||
Db        947 LASVFSLVWVKIDPFVSKGDSNL-TQGCIAIDC 978


RESULT 7
Q9SN37
ID   Q9SN37      PRELIMINARY;      PRT;   958 AA.
AC   Q9SN37;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Cellulose synthase-like protein.
GN   F28A21.190 OR AT4G18780.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Bevan M., Mueller M.W., Muendlein A., Felber R., Bancroft I.,
RA   Mewes H.W., Mayer K.F.X., Lemcke K., Schueller C.;
RL   Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
```

```
A   EU Arabidopsis sequencing project;
L   Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
N   [3]
P   SEQUENCE FROM N.A.
A   Mueller M.W., Muendlein A., Felber R., Mewes H.W., Lemcke K.,
A   Mayer K.F.X.;
L   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
N   [4]
P   SEQUENCE FROM N.A.
A   EU Arabidopsis sequencing project;
L   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
R   EMBL; AL035526; CAB37463.1; -.
R   EMBL; AL161549; CAB78880.1; -.
R   InterPro; IPR005150; Cellulose_synt.
R   Pfam; PF03552; Cellulose_synt; 1.
Q   SEQUENCE   958 AA;  108520 MW;  B3836EF4BC697502 CRC64;

 Query Match             80.2%;  Score 4202;  DB 10;  Length 958;
 Best Local Similarity   79.1%;  Pred. No. 0;
 Matches  784;  Conservative   85;  Mismatches   76;  Indels   46;  Gaps    9;

Qy          1 MMESGAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEINEGRKVCLRCGSPYDE 60
              |||| :|||:|||||::|  :||| ||||||||:|:||:| |:|  :|||::|||||:|||| 
Db          1 MMESRSPICNTCGEEIGVKSNGEFFVACHECSFPICKACLEYEFKEGRRICLRCGNPYDE 60

Qy         61 NLLDDVEKKGSGNQSTMASHLNDSQDVGIHARHISSVSTVDSEMNDEYGNPIWKNRVKSC 120
              |: |||| | |  ||          | |||||||:|||:|||:|||||||||||||:| 
Db         61 NVFDDVETKTSKTQSI---------DSGIHARHISTVSTIDSELNDEYGNPIWKNRVESW 111

Qy        121 ---KDKENKKKKRSP---KAETEPAQVPTEQQMEEKP---SAEASEPLSIVYPIPRNKLT 171
                 |||::||||: |   |||   ||:||:| ||: |    : |:: ||:| |||| |:| 
Db        112 KDKKDKKSKKKKKDPKATKAEQHEAQIPTQQHMEDTPPNTESGATDVLSVVIPIPRTKIT 171

Qy        172 PYRAVIIMRLVILGLFFHFRITNPVDSAFGLWLTSVICEIWFAFSWVLDQFPKWNPVNRE 231
              || |||||:|| |||::|||:|||||:|||||||||||||| ||||||||:|:|||| 
Db        172 SYRIVIIMRLIILALFFNYRITHPVDSAYGLWLTSVICEIWFAVSWVLDQFPKWSPINRE 231

Qy        232 TYIERLSARYEREGEPSQLAGVDFFVSTVDPLKEPPLITANTVLSILAVDYPVDKVSCYV 291
              |||:|||||:||||| |||| ||||||||||||||||||||||||||:|||||||||| 
Db        232 TYIDRLSARFEREGEQSQLAAVDFFVSTVDPLKEPPLITANTVLSILALDYPVDKVSCYV 291

Qy        292 SDDGAAMLSFESLVETAEFARKWVPFCKKFSIEPRAPEFYFSQKIDYLKDKVQPSFVKER 351
              |||||||||||||||||:|||||||||||:|||||||||||| |||||:|||||||||| 
Db        292 SDDGAAMLSFESLVETADFARKWVPFCKKYSIEPRAPEFYFSLKIDYLRDKVQPSFVKER 351

Qy        352 RAMKRDYEEYKVRVNALVAKAQKTPEEGWTMQDGTPWPGNNTRDHPGHDSGLPWEIL--- 408
              |||||||||:|:|:||||||||||||||||||||||| ||||||||||    |:    | 
Db        352 RAMKRDYEEFKIRMNALVAKAQKTPEEGWTMQDGTSWPGNNTRDHP----GMIQVFLGYS 407

Qy        409 GARDIEGNELPRLVYVSREKRPGYQHHKKAGAENALVRVSAVLTNAPYILNVDCDHYVNN 468
              |||||||||||||||||||||||||||||||||||||||||||||||:|||:|||||||| 
Db        408 GARDIEGNELPRLVYVSREKRPGYQHHKKAGAENALVRVSAVLTNAPFILNLDCDHYVNN 467

Qy        469 SKAVREAMCILMDPQVGRDVCYVQFPQRFDGIDKSDRYANRNVVFFDVNMKGLDGIQGPV 528
              ||||||||| |||| ||:|||:|||||||||||||||||||:|||||||:|||||||| 
Db        468 SKAVREAMCFLMDPVVGQDVCFVQFPQRFDGIDKSDRYANRNIVFFDVNMRGLDGIQGPV 527

Qy        529 YVGTGCVFNRQALYGYGPPSMPSLRKRKDSSSCFSCCCPSKKKPAQDPAEVYRDAKREDL 588
              ||||| || ||||||| ||| | :  :  ||   |||| :|||  |||:|:|:|||||:| 
Db        528 YVGTGTVFRRQALYGYSPPSKPRILPQSSSS---SCCCLTKKKQPQDPSEIYKDAKREEL 584

Qy        589 NAAIFNLTEIDNYDEHERSMLISQLSFEKTFGLSSVFIESTLMENGGVPESANSPPFIKE 648
              :|||||| ::|||||::|||||||| |||||||||:|||||||||||||:| |    ||| 
Db        585 DAAIFNLGDLDNYDEYDRSMLISQTSFEKTFGLSTVFIESTLMENGGVPDSVNPSTLIKE 644

Qy        649 AIQVIGCGYEEKTEWGKQIGWIYGSVTEDILSGFKMHCRGWRSIYCMPVRPAFKGSAPIN 708
              || || |||||||||||:|||||||:|||||:|||||||||||||||||:|||||||||| 
Db        645 AIHVISCGYEEKTEWGKEIGWIYGSITEDILTGFKMHCRGWRSIYCMPLRPAFKGSAPIN 704

Qy        709 LSDRLHQVLRWALGSVEIFFSRHCPLWYGFGGGRLKWLQRLAYINTIVYPFTSLPLIAYC 768
              |||||||||||||||||| |||||||||  ||||| ||||||||||||||||||:||| 
Db        705 LSDRLHQVLRWALGSVEIFLSRHCPLWYGCSGGRLKLLQRLAYINTIVYPFTSLPLVAYC 764

Qy        769 TIPAVCLLTGKFIIPTLSNLASMLFLGLFISIIVTAVLELRWSGVSIEDLWRNEQFWVIG 828
              |:| :||| |||  ||:|||| '                    |||||:|||||.,||| 
Db        765 TLPAICLLTGKFIIPTLSNLASMLFL--  -  -  --------GVSIEDLWRNEQFWVIG 817

Qy        829 GVSAHLFAVFQGFLKMLAGLDTNFTVTAKAADEAEFGELYMVKWTTLLIPPTTLLIINMS 888
              ||||||||||||:|||:|:|:|||||:|  |:  |||||:|||||||||||:|||| : 
Db        808 GVSAHLFAVFQGFLKMLAGLDTNFTVTSKTADDDEFGELYIVKWTTLLIPPTSLLIINLV 867

Qy        889 G-CAGFSDALNKGYAWGPLFGKVFFAFWVILHLYPFLKGLMGRQNLTPTIVVLWSVLLA 947
              | |||||||||:|   ||| :|||:'|||: |||||||||||| |||.:|||: .| 
Db        868 GVVAGFSDALNKGYAWGPLFGKVFFAFWVILHLYPFLKGLMGRQNRTPTIVILWSILLA 927

Qy        948 SVFSLVWVKINPFVAKVDNTLVAETCISIDC 978
              ||| ||||:|||| : |  :: :: || 
Db        928 SVFSLVWVRINPFVKKTDTTSLSLNCLLIDC 958
```

```
RESULT 8
Q943H3
ID   Q943H3      PRELIMINARY;      PRT;   939 AA.
AC   Q943H3;
DT   01-DEC-2001 (TrEMBLrel. 14, Created)
DT   01-DEC-2001 (TrEMBLrel. 14, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Putative cellulose synthase.
GN   P0046E05.5.
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=4530;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Nipponbare;
RA   Sasaki T., Matsumoto T., Yamamoto K.;
RT   "Oryza sativa nipponbare (GA) genomic DNA, chromosome 1, PAC
RT   clone:P0046E05.";
RL   Submitted (FEB 2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AP003237; BAB67906.1; -.
DR   Gramene; Q943H3; -.
DR   InterPro; IPR005150; Cellulose_synt.
DR   Pfam; PF03552; Cellulose_synt; 1.
SQ   SEQUENCE   939 AA;  10[illegible] MW;  49973496B712D86B CRC64;

  Query Match             7[illegible] 6%;  Score 3699;  DB 10;  Length 939;
  Best Local Similarity   7[illegible] 6%;  Pred. No. 2.2e-283;
  Matches  707;  Conservative   83;  Mismatches  125;  Indels   86;  Gaps   12;

Qy          2 MESGAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEINEGRKVCLRCGSPY--- 58
              ||:| | :  ||:    ||:    ||  :|:|:| : :  |||  | |||  | 
Db          1 MESGVPPCAACGD  --DAH-  --AACRACSYALCKACLDEDAAEGRTTCARCGGEYGAP 52

Qy         59 ---------DENLLDDVEKKGSG--NQSTMASHLNDSQDVGI-------HARHISSVST 99
                      :| : :  |    |   : || ! |:| || |:        ||| |:|| 
Db         53 DPAHGQGAVVREEVESHEPAAGGVRERVTMASQLSDHQDEGVHARTMSTHARTISSVSG 112

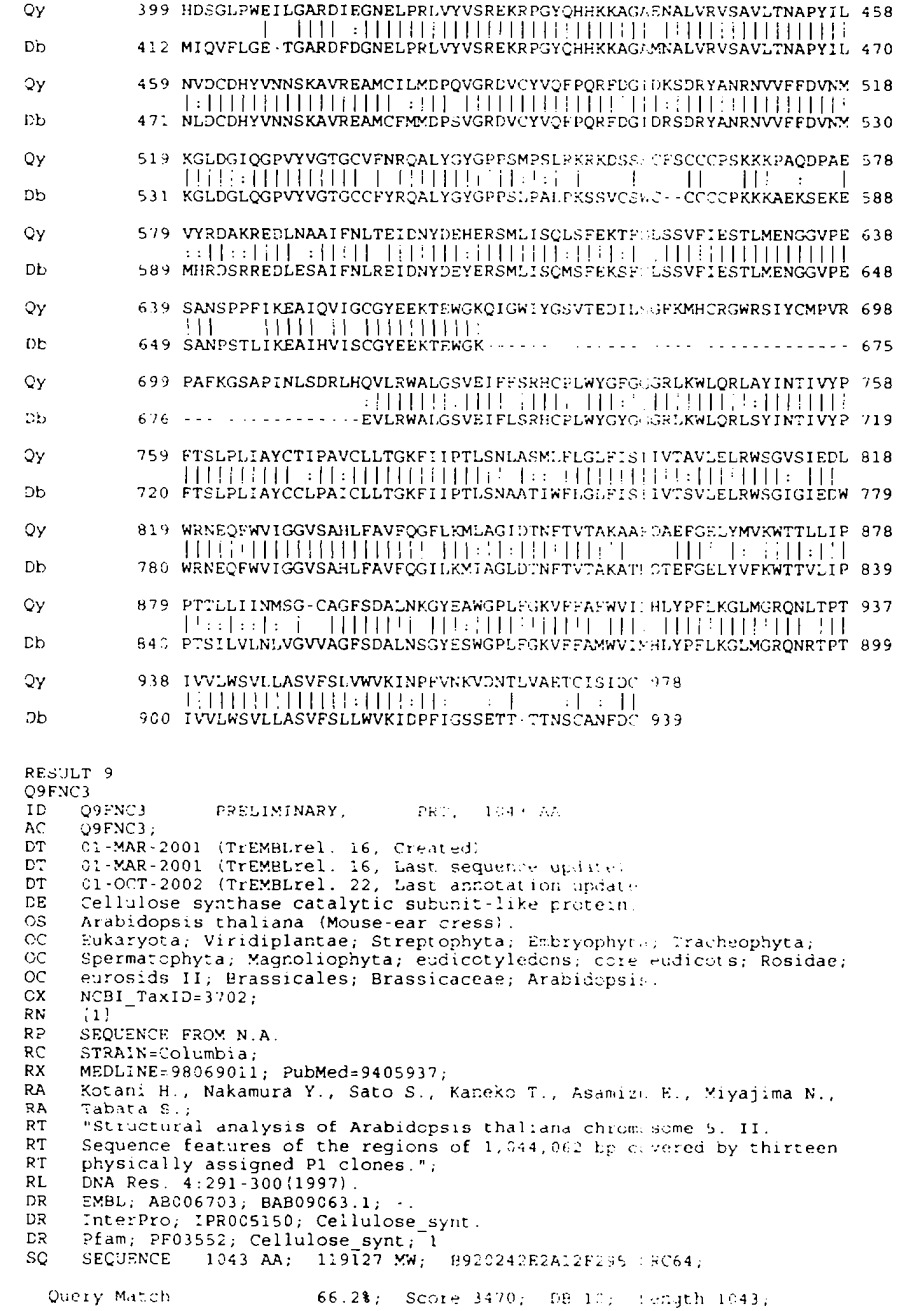
Qy        100 VDSEMNDEYGNPIWKNRVKSCKDKENKKKKRSPKAETEPAQVPTEQQ-MEEKPSAEASEP 158
              | |:||| , ||: ||:|  |:|:  :'|  :  '  :    | |:| |:||   :|  | 
Db        113 VGSELNDESGKPIWKNRVESWKEKKKEKKASAKKAAAKAQAPPVEEQIMDEKDLTDAYEP 172

Qy        159 LSIVYPIPRNKLTPYRAVIIMRLVILGLFFHFRITNPVDSAFGLWLTSVICEIWFAFSWV 218
              || : || :|||||||| |||| |:|:|:|||:|||||| ||||||:||||:|||  ||: 
Db        173 LSRIIPISKNKLTPYRAVIIMRLVVLGLFFHYRITNPVYSAFGLWMTSVICEIWFGFSWI 232

Qy        219 LDQFPKWNPVNRETYIERLSARYEREGEPSQLAGVDFFVSTVDPLKEPPLITANTVLSIL 278
              ||||||| |:||||||:|| |||  :|, |  ' |||||||||||||,'||||,'||||
Db        233 LDQFPKWCPINRETYVDRLIARY-GDGEDSGLAPVDFFVSTVDPLKEPPLITANTVLSIL 291

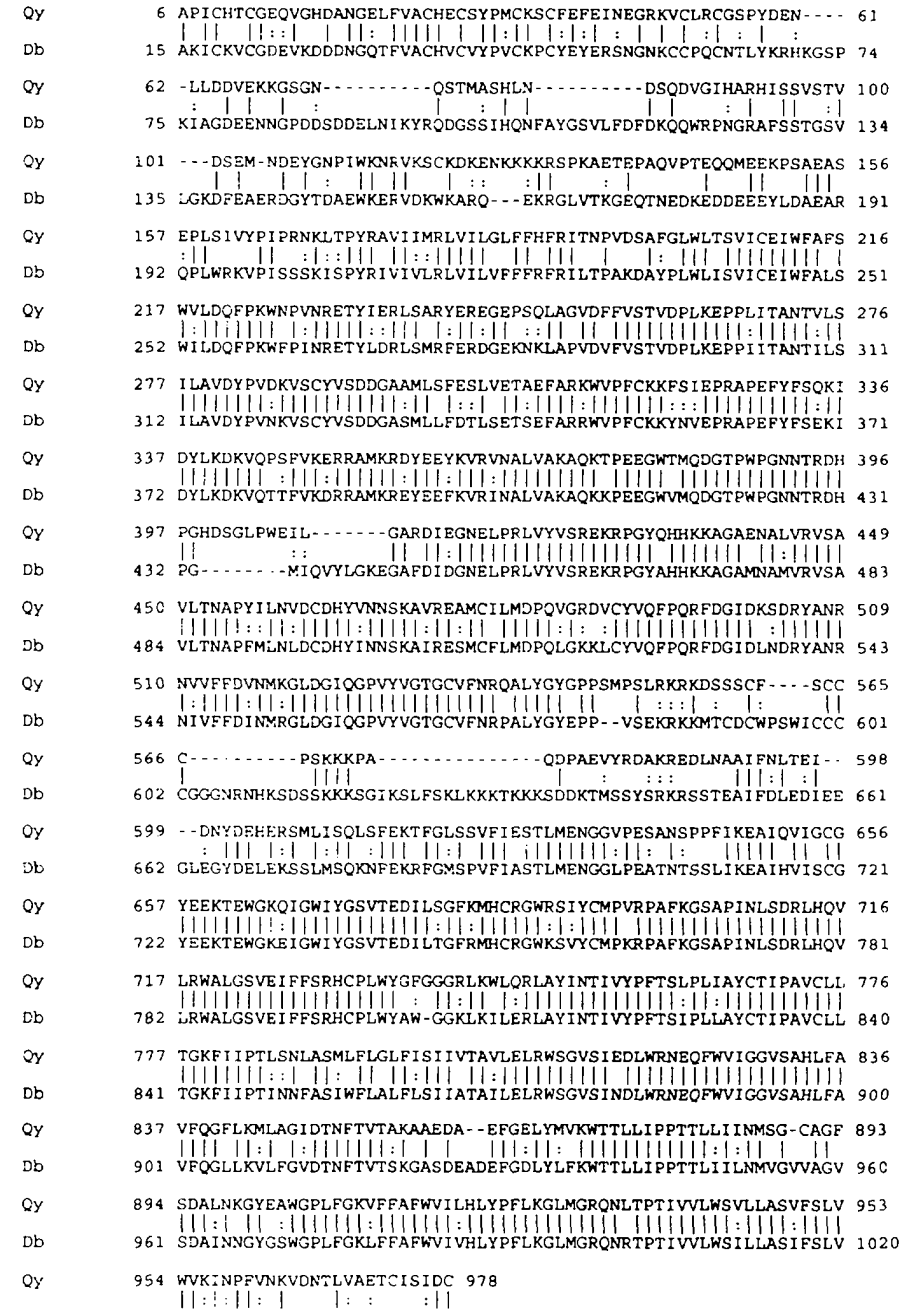
Qy        279 AVDYPVDKVSCYVSDDGAAMLSFESLVETAEFARKWVPFCKKFSIEPRAPEFYFSQKIDY 338
              |||||||:|||||.'|:|||:|||| ||||:||:|||||||:|||||||||||||||||| 
Db        292 AVDYPVEKISCYVSDDGSAMLTFESLAETAEFARRWVPFCKKYSIEPRAPEFYFSQKIDY 351

Qy        339 LKDKVQPSFVKERRAMKRDYEEYKVRVNALVAKAQKTPEEGWTMQDGTPWPGNNTRDHPG 398
```



RESULT 9
Q9FNC3
ID Q9FNC3 PRELIMINARY; PRT; 1044 AA.
AC Q9FNC3;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Cellulose synthase catalytic subunit-like protein.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Columbia;
RX MEDLINE=98069011; PubMed=9405937;
RA Kotani H., Nakamura Y., Sato S., Kaneko T., Asamizu E., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. II.
RT Sequence features of the regions of 1,044,062 bp covered by thirteen
RT physically assigned P1 clones.";
RL DNA Res. 4:291-300(1997).
DR EMBL; AB006703; BAB09063.1; -.
DR InterPro; IPR005150; Cellulose_synt.
DR Pfam; PF03552; Cellulose_synt; 1
SQ SEQUENCE 1043 AA; 119127 MW; B920242E2A12F295 CRC64;

Query Match 66.2%; Score 3470; DB 1; Length 1043;



```
b       1021 WVRIDPFLPKQTGPLLKQ--CGVDC 1043


ESULT 10
8GZN8
D   Q8GZN8      PRELIMINARY;      PRT;  1055 AA.
C   Q8GZN8;
T   01-MAR-2003 (TrEMBLrel. 23, Created)
T   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
T   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
E   Cellulose synthase.
N   IRX5.
S   Arabidopsis thaliana (Mouse-ear cress).
C   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
C   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
C   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
X   NCBI_TaxID=3702;
N   [1]
P   SEQUENCE FROM N.A.
A   Taylor N.G., Vickers K., Turner S.R.;
T   "IRX5 encodes a cellulose synthase.";
L   Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
R   EMBL; AF458083; AAO15532.1; -.
Q   SEQUENCE   1055 AA;  120108 MW;  9C2FC0DD289D4682 CRC64;

 Query Match             66.1%;  Score 3464.5;  DB 10;  Length 1055;
 Best Local Similarity   63.1%;  Pred. No. 9.1e-265;
 Matches  663;  Conservative  122;  Mismatches  173;  Indels   93;  Gaps   16;

y          6 APICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEINEGRKVCLRC--------GSP 57
             |  ||  ||::|  | ||: |||||  | ||:|| |:|:| : | | | :|        |||
b         20 AKICKVCGDEVKDDDNGQTFVACHVCVYPVCKPCYEYERSNGNKCCPQCNTLYKRHKGSP 79

y         58 ------------YDENLLDDVEKKGSGNQSTMA-----SHLNDSQDVGIHARHISSVST 99
                        |: |     : ||      |        |  |     : |  ||  :
b         80 KIAGDEENNGPDDSDDELNIKYRQDGSSIHQNFAYGSENGDYNSKQQCRPNGRAFSSTGS 139

y        100 V---DSEM-NDEYGNPIWKNRVKSCKDKENKKKKRSPKAETEPAQVPTEQQMEEKPSAEA 155
             |   | |    | | :  || ||    | :: |:    :   :|    :    |:  ||   |||
b        140 VLGKDFEAERDGYTDAEWKERVDKWKARQEKRGLVTKGEQTNEDKEDDEE--EELLDAEA 197

y        156 SEPLSIVYPIPRNKLTPYRAVIIMRLVILGLFFHFRITNPVDSAFGLWLTSVICEIWFAF 215
              :||    ||  :|::||| ||::||||  || |||  |   |: ||| ||||||||
b        198 RQPLWRKVPISSSKISPYRIVIVLRLVILVFFFRFRILTPAKDAYPLWLISVICEIWFAL 257

y        216 SWVLDQFPKWNPVNRETYIERLSARYEREGEPSQLAGVDFFVSTVDPLKEPPLITANTVL 275
             ||:|||||| |:|||||::||| |:||:|| ::|| || |||||||||||:|||||:|
b        258 SWILDQPPKWFPINRETYLDRLSMRFERDGEKNKLAPVDVFVSTVDPLKEPPIITANTIL 317

y        276 SILAVDYPVDKVSCYVSDDGAAMLSFESLVETAEFARKWVPFCKKFSIEPRAPEFYFSQK 335
             |||||||||:||||||||||:|| |::| ||:||||:|||||||:::|||||||||:|
b        318 SILAVDYPVNKVSCYVSDDGASMLLFDTLSETSEFARRWVPFCKKYNVEPRAPEFYFSEK 377

y        336 IDYLKDKVQPSFVKERRAMKRDYEEYKVRVNALVAKAQKTPEEGWTMQDGTPWPGNNTRD 395
             ||||||||| :|||:||||||:|||:|||:|||||||||| ||||| |||||||||||||
b        378 IDYLKDKVQTTFVKDRRAMKREYEEFKVRINALVAKAQKKPEEGWVMQDGTPWPGNNTRD 437

y        396 HPGHDSGLPWEIL-------GARDIEGNELPRLVYVSREKRPGYQHHKKAGAENALVRVS 448
             |||         ::       || ||:|||||||||||||||||| |||||| ||:||||
b        438 HPG--------MIQVYLGKEGAFDIDGNELPRLVYVSREKRPGYAHHKKAGAMNAMVRVS 489

y        449 AVLTNAPYILNVDCDHYVNNSKAVREAMCILMDPQVGRDVCYVQFPQRFDGIDKSDRYAN 508
             |||||||::||:|||||:|||||:||:|| |||||:|: :|||||||||||| :|||||
b        490 AVLTNAPFMLNLDCDHYINNSKAIRESMCFLMDPQLGKKLCYVQFPQRFDGIDHNDRYAN 549

y        509 RNVVFFDVNMKGLDGIQGPVYVGTGCVFNRQALYGYGPPSMPSLRKRKDSSSCF----SC 564
             ||:||||:||:|||||||||||||||||| ||||| ||   | :::| :  |:      |
b        550 RNIVFFDINMRGLDGIQGPVYVGTGCVFNRPALYGYEPP--VSEKRKKMTCDCWPSWICC 607

y        565 CC-----------------------------PSKKKPAQDPAEVYRDAKREDLNAAIFN 594
             ||                             | |   |   :   :::    |||:
b        608 CCGGGGNRNHHKSKSSDSSSKKKSGIKSLLSKLKKKNKKKSDDKTMSSYSRKRSATEAIFD 667
```

```
Qy       595 LTEI----ENYDEHRKSMLISQLSFEKTFGLSSVFIESTLMENGGVPESANSPPFIKEAI 650
             | :|    : ||| |:|  :|| :||| ||:| ||| |||||||:||: |:   ||||
Db       668 LEDIEEGLEGYDEIRKSSLMSQKNFEKRFGMSPVFIASTLMENGGLPEATNTSSLIKEAI 727

Qy       651 QVIGCGYEEKTEWGKQIGWIYGSVTEDILSGFKMHCRGWRSIYCMPVRPAFKGSAPINLS 710
             || ||:|||||||:|:|||:|||||:||:|||||:|:|||| ||||:|||||'||
Db       728 HVISCGYEEKTEWGKEIGWIYGSVTEDILTGFRMHCRGWKSVYCMPKRPAFKGSAPINLS 787

Qy       711 DRLHQVLRWALGSVEIFFSRHCPLWYGFGGGRLKWLQRLAYINTIVYPFTSLPLIAYCTI 770
             |||||||||||:|||,,|||:|||| : ||:|. |:|||||||||||||,:||:|:|||
Db       788 DRLHQVLRWALGSVEIFFSRHCPLWYAW-GGKLKILERLAYINTIVYPFTSIPLLAYCTI 846

Qy       771 PAVCLLTGKFIIPT.SNLASMLFLGLFISIIVTAVLELRWSGVSIEDLWRNEQFWVIGGV 830
             ||||||||||||:| |:  || |:||| ||:|||||||||| ||||||||||:|||||||
Db       847 PAVCLLTGKFIIPT.NNFASIWFLALFLSIIATAILELRWSGVSINDLWRNEQFWVIGGV 906

Qy       831 SAHLFAVFQGFLKM.AGIDTNFTVTAKAAEDA--EFGELYMVKWTTLLIPPTTLLIINMS 888
             ||:|||||:| |  |:|||||:|| | :  |||:||: |||||||:|||:|:.|
Db       907 SAHLFAVFQGLLP  FGVDTNFTVTSKGASDEADEFGDLYLFKWTTLLIPPTTLIIINMV 966

Qy       889 GVAGFSDAINN  AWGPLFGKLFFAFWVIMHLYPFLKGLMGRQNRTPTIVVLWSVLLA 947
             |  ||   :|    :  |||  :||  |:| |||||| |||| :|||||:|:|||
Db       967 GVVAGVSDAINN  RWGRLFGKLFFAFWVIMHLYPFLKGLMGRQNRTPTIVVLWSILLA 1026

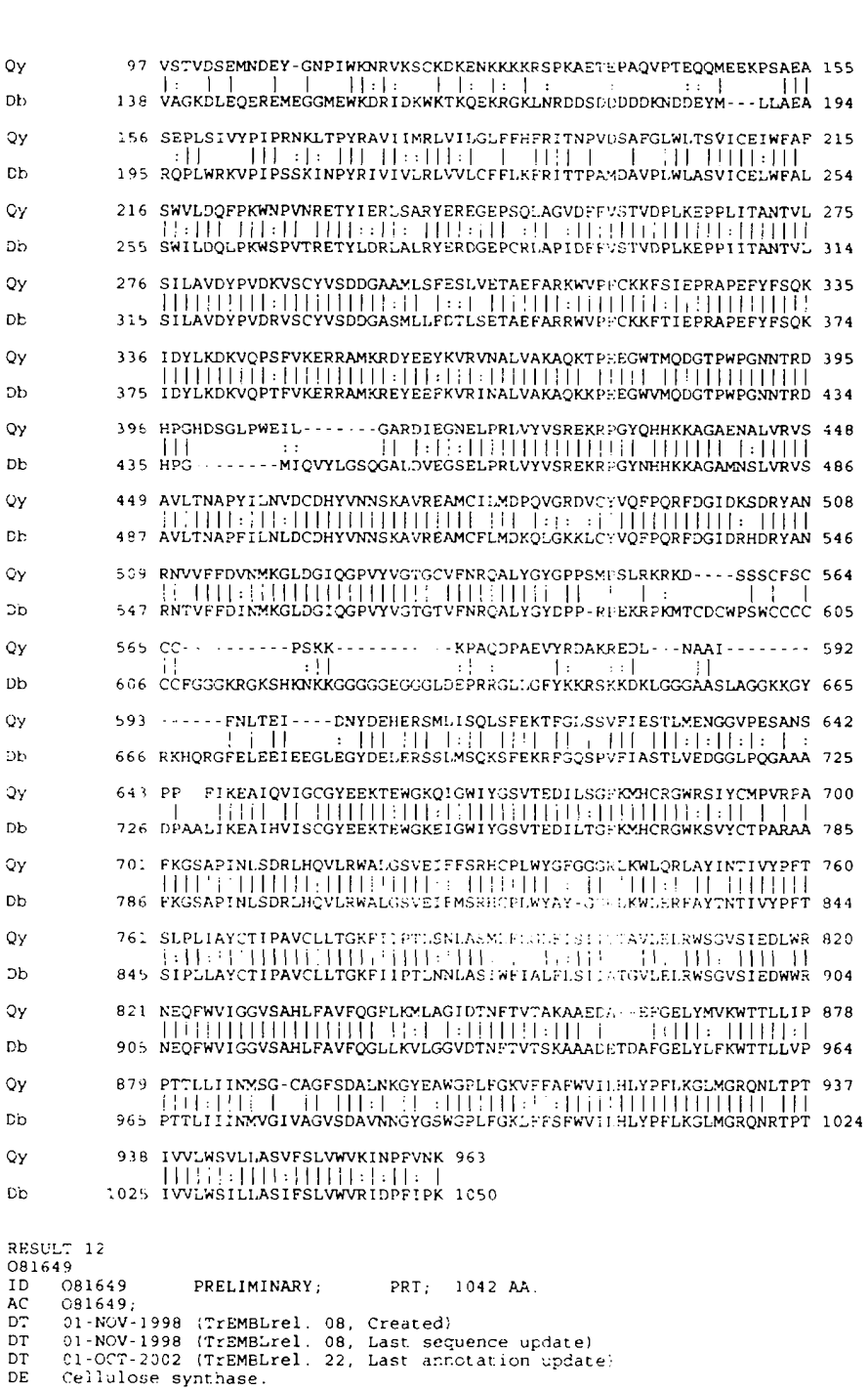
Qy       948 SVFSLVWVKINPFV.NKVDNTLVAETCISIDC 978
             |: |||:.:|:::  |    |: :   : |
Db      1027 SIFSLVWVRIDPFI PKQTGPLLKQ- CGVDC 1055
```

```
RESULT 11
Q9AV71
ID   Q9AV71      PRELIMINARY;      PRT;  1063 AA.
AC   Q9AV71;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Putative cellulose synthase.
GN   OSJNBA0006L06.10.
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza
OX   NCBI_TaxID=4530;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Nipponbare;
RA   Buell C.R., Yuan Q., Ouyang S., Moffat K.S., Hill J.N., Burr P.C.,
RA   Hsiao J., Zismann V., Pai G., Bowman C.L., Fujii C.Y., VanAken S.E.,
RA   Bowman C.L., Craven B., Utterback T.R., Khalak H., Feldblyum T.V.,
RA   Quackenbush J., White O., Salzberg S.L., Fraser C.M.;
RT   "Oryza sativa chromosome 10 BAC OSJNBa0006L06 genomic sequence.";
RL   Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC022457; AAK27814.1; -.
DR   Gramene; Q9AV71; -.
DR   InterPro; IPR005150; Cellulose_synt.
DR   InterPro; IPR001841; Zf_ring.
DR   Pfam; PF03552; Cellulose_synt; 1.
DR   SMART; SM00184; RING; 1.
DR   PROSITE; PS50089; ZF_RING_2; 1.
SQ   SEQUENCE   1063 AA;  119787 MW;  5E02386EFAECDE4E CRC64;

  Query Match             6[illegible] 0%;  Score 3406;  DB 10;  Length 1063;
  Best Local Similarity   6[illegible].2%;  Pred. No. 3.9e-260;
  Matches  661;  Conservative  109;  Mismatches  172;  Indels  104;  Gaps   17;

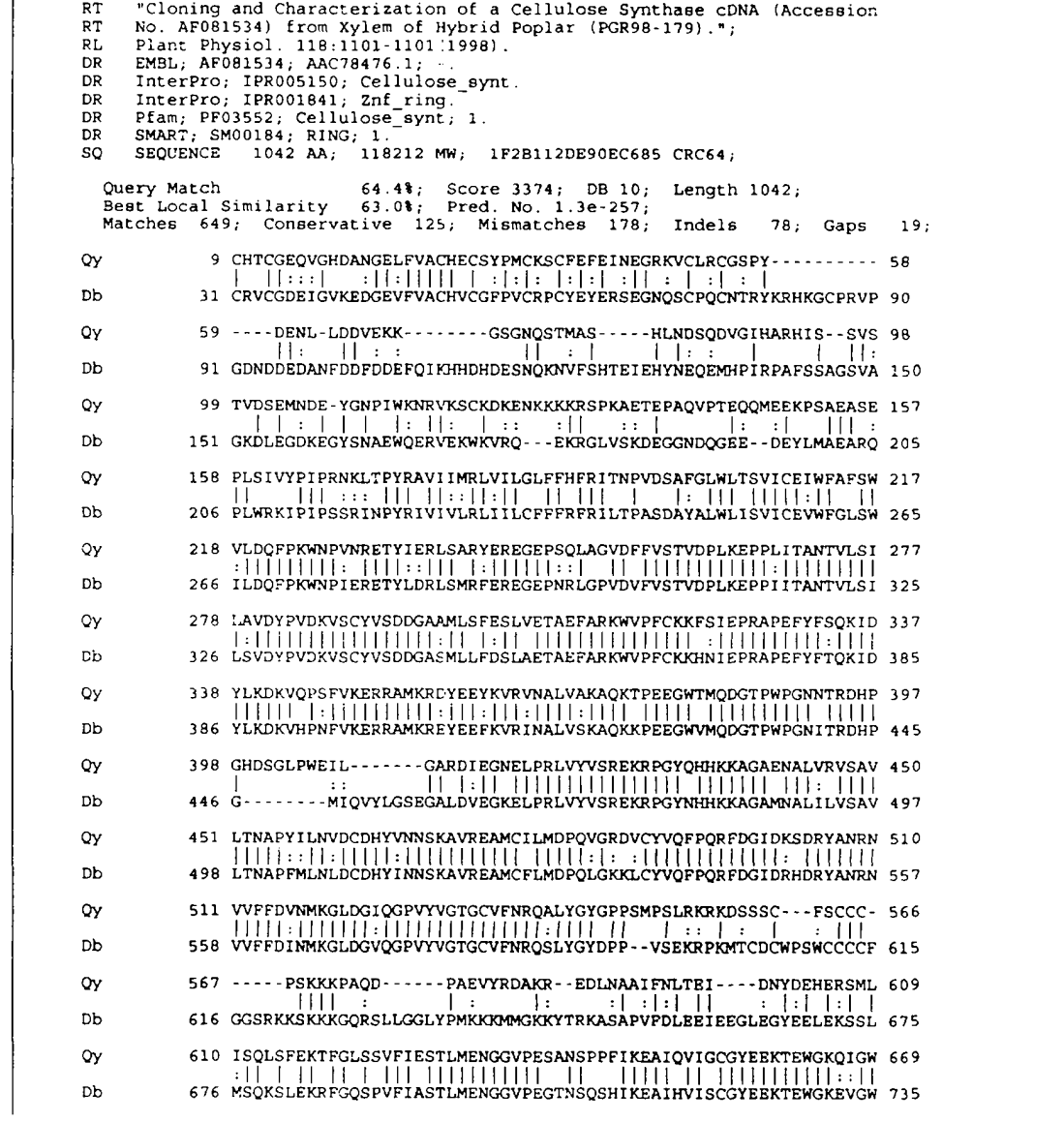
Qy         9 CHTCGEQVGHDAN ELFVACHECSYPMCKSCFEFEINEGRKVCLRCGSPY---------- 58
             |  |||:|    :|: |||| || ::|| |:|:| :|| : | :| : |
Db        18 CRVCGEEVAARED KPFVACAECGFPVCKPCYEYERSEGTQCCPQCNTRYKRHKGCPRVE 77

Qy        59 ----DENLLDDVEKKGSGNQSTMASHLNDSQDVGIHARH------- ---- ---- ISS 96
                 |   :|| |::      |      :: :  ::: :                  |
Db        78 GDEDDGGDMDDFEFEFQIKSPTKQKPPHEPVNFDVYSENGEQPAQKWRPGGPALSSFTGS 137
```



RESULT 12
O81649
ID O81649 PRELIMINARY; PRT; 1042 AA.
AC O81649;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Cellulose synthase.
GN CEL1.
OS Populus x canescens.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Malpighiales; Salicaceae; Populus.
OX NCBI_TaxID=80863;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=717; TISSUE=Xylem;
RX MEDLINE=99105148; PubMed=9890811;
RA Wang H., Loopstra C.A.;
RT "Cloning and Characterization of a Cellulose Synthase cDNA (Accession
RT No. AF081534) from Xylem of Hybrid Poplar (PGR98-179).";
RL Plant Physiol. 118:1101-1101(1998).
DR EMBL; AF081534; AAC78476.1; -.
DR InterPro; IPR005150; Cellulose_synt.
DR InterPro; IPR001841; Znf_ring.
DR Pfam; PF03552; Cellulose_synt; 1.
DR SMART; SM00184; RING; 1.
SQ SEQUENCE 1042 AA; 118212 MW; 1F2B112DE90EC685 CRC64;

Query Match 64.4%; Score 3374; DB 10; Length 1042;
Best Local Similarity 63.0%; Pred. No. 1.3e-257;
Matches 649; Conservative 125; Mismatches 178; Indels 78; Gaps 19;



```
y       670 IYGSVTEDILSGFKMHCRGWRSIYCMPVRPAFKGSAPINLSDRLHQVLRWALGSVEIFFS 729
            ||||||||||:|||||||||||:|| | |||||||||||||||||||||||||:||| |
b       736 IYGSVTEDILTGFKMHCRGWRSVYCSPKRPAFKGSAPINLSDRLHQVLRWALGSIEIFLS 795

y       730 RHCPLWYGFGGGRLKWLQRLAYINTIVYPFTSLPLIAYCTIPAVCLLTGKFIIPTLSNLA 789
            |||||||: ||:|| |:|||||||||||||||:||:|||| |||||||||||||||:|||
b       796 HHCPLWYGY-GGKLKLLERLAYINTIVYPFTSIPLLAYCTTPAVCLLTGKFIIPTLNNLA 854

y       790 SMLFLGLFISIIVTAVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQGFLKMLAGID 849
            |: | | |       : : ||||||:|| |||||||||||| |||||||| |:| |:|
b       855 SIWFPGPFHLNHSNICVGVGWSGVSIQDLGRNEQFWVIGGVSGHLFAVFQGFFKVLGGVD 914

y       850 TNFTVTAKAAEDAEFGELYMVKWTTLLIPPTTLLIINMSG-CAGFSDALNKGYEAWGPLF 908
            ||||||:|:|:||||||||: ||||||||||||:|:|| |  || |  :| || :|||||
b       915 TNFTVTSKSADDAEFGELYLFKWTTLLIPPTTLIILNMVGVVAGVSARINNGYGSWGPLF 974

y       909 GKVFFAFWVILHLYPFLKGLMGRQNLTPTIVVLWSVLLASVFSLVWVKINPFVNKVDNTL 968
            ||:|||||||:|||||||||||:|| ||||||||||||||:|||:||:|:||: |    :
b       975 GKLFFAFWVIVHLYPFLKGLMGKQNRTPTIVVLWSVLLASIFSLIWVRIDPFLPKQTGPI 1034

y       969 VAETCISIDC 978
            : :    ::|
b      1035 LKQ--CGVEC 1042


ESULT 13
9LLI1
D   Q9LLI1      PRELIMINARY;      PRT;  1079 AA.
C   Q9LLI1;
T   01-OCT-2000 (TrEMBLrel. 15, Created)
T   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
T   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
E   Cellulose synthase-9.
N   CESA-9.
S   Zea mays (Maize).
C   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
C   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
C   PACCAD clade; Panicoideae; Andropogoneae; Zea.
K   NCBI_TaxID=4577;
N   [1]
P   SEQUENCE FROM N.A.
X   MEDLINE=20398328; PubMed=10938350;
A   Holland N., Holland D., Helentjaris T., Dhugga K.S.,
A   Xoconostle-Cazares B., Delmer D.P.;
T   "A comparative analysis of the plant cellulose synthase (CesA) gene
T   family.";
L   Plant Physiol. 123:1313-1324(2000).
R   EMBL; AF200533; AAF89969.1; -.
R   InterPro; IPR005150; Cellulose_synt.
R   InterPro; IPR001841; Znf_ring.
R   Pfam; PF03552; Cellulose_synt; 1.
R   SMART; SM00184; RING; 1.
D   SEQUENCE   1079 AA;  120762 MW;  EB334272C2ED2DD7 CRC64;

 Query Match             64.2%;  Score 3366.5;  DB 10;  Length 1079;
 Best Local Similarity   59.6%;  Pred. No. 5.3e-257;
 Matches  645;  Conservative  133;  Mismatches  178;  Indels  127;  Gaps   19;

/         5 GAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEINEGRKVCLRC--------GS 56
            |  :|  ||: ||  | |::| ||  | :|:|: |:|:|  :| : | :|        ||
>        15 GGQVCQICGDGVGTTAEGDVFTACDVCGFPVCRPCYEYERKDGTQACPQCKNKYKRHKGS 74

/        57 P------------------YDENLLDDVEKKGSGNQSTMASHLNDSQDVG--------- 88
            |                  |  :  || ::| :    :   :   | |||
>        75 PAIRGEEGDDTDADDASDFNYPASGNDDQKQKIADRMRSWRMNAGGSGDVGRPKYDSGEI 134

/        89 ---------IHARHISSV-----------------------------------STVDS 102
                     |   :| ||                                   |
>       135 GLTKYDSGEIPRGYIPSVTNSQISGEIPGASPDHMMMSPTGNIGRRAPFPYMNHSSNPSR 194

/       103 EMNDEYGNPIWKNRVKSCKDKENK----KKKRSPKAETEPAQV-----PTEQQMEEK-PS 152
            | :   ||  || ||   | |::|       : | :|   |      |:  ||:   :
>       195 EFSGSVGNVAWKERVDGWKMKQDKGTIPMTNGTSIAPSEGRGVGDIDASTDYNMEDALLN 254

Qy      153 AEASEPLSIVYPIPRNKLTPYRAVIIMRLVILGLFFHFRITNPVDSAFGLWLTSVICEIW 212
                |  :|||   |:| ::: ||: ||::||:: :| |:|||||  :|: |:  |||||||
Db      255 DETRQPLSRKVPLPSSRINPYRMVIVLRLIVLSIFLHYRITNPVRNAYPLWLLSVICEIW 314

Qy      213 FAFSWVLDQFPKWNPVNRETYIERLSARYEREGEPSQLAGVDFFVSTVDPLKEPPLITAN 272
            || ||:|||:||| ::||||:::|:: ||:|'|||||||| || |||||||:|||||:|||
Db      315 FALSWILDQFPKWFPINRETYLDRLALRYDREGEPSQLAAVDIFVSTVDPMKEPPLVTAN 374

Qy      273 TVLSILAVDYPVDKVSCYVSDDGAAMLSFESLVETAEFARKWVPFCKKFSIEPRAPEFYF 332
            |||||||:|||||'|||:|||||:|:|::| ||:|||||||| ||::|||||:|:||
Db      375 TVLSILAVDYPVDKVSCYVSDDGAAMLTFDALAETSEFARKWVPFVKKYNIEPRAPEWYF 434

Qy      333 SQKIDYLKDKVQPSFVKERRAMKREYEEYKVRVNALVAKAQKTPEEGWTMQDGTPWPGNN 392
            |||||||||| ! |||:|||||:|||:|:||| |||||| |||| ||||||||||
Db      435 SQKIDYLKDKVHPSFVKDRRAMKREYEEFKIRVNGLVAKAQKVPEEGWIMQDGTPWPGNN 494

Qy      393 TRDHP-------GHDSGLPWEIIGARDIEGNELPRLVYVSREKRPGYQHHKKAGAENALV 445
            |: ||        :  :!         | :|||||:|||||'|||:|:|||||: ||||
Db      495 TRDHPGMIQVFLGH-SGL         DTEGNELPRLVYVSREKRPGFQHHKKAGAMNALV 546

Qy      446 RVSAVLTNAPYILNLDCDHYVNNSKAVREAMCFLMDPQVGRDVCYVQFPQRFDGIDKSDR 505
            |||||:  |:|:|: ||:|:|||:|||  :|| :|| |||| |||||:|||:||:||
Db      547 RVSAVLTNGQYMLNLDCDHYINNSKALREAMCFLMDPNLGRSVCYVQFPQRFDGIDRNDR 606

Qy      506 YANRNVVFFDVNMKGLDGIQGPVYVGTGCVFNRQALYGYGPPSMPSLRKRKDSSSCFSCC 565
            ||||| ||||:|:::|||||||:|||||||||| ||||| ||    ::::|      |
Db      607 YANRNTVFFDINLRGLDGIQGPVYVGTGCVFNRTALYGYEPP----IKQKK--GGPLSSL 660

Qy      566 CPSKKKPAQDP--AEVYRDAKREDLNAAIFNLTEID------NYDEHERSMLISQLSFEK 617
            |  :|| ::    ::  :  !  | :  :||| :|:       :|: |:|:|:||:| ||
Db      661 CGGRKKGSKSKKGSDKKKSQKHVDSSVPVFNLEDIEEGVEGAGFDD-EKSLLMSQMSLEK 719

Qy      618 TFGLSSVFIESTLMENGGVPESANSPPFIKEAIQVIGCGYEEKTEWGKQIGWIYGSVTED 677
            || |: |: ||:|' ||||:||      :|||| |: |||| :|:|||:|||| :||||:||
Db      720 RFGQSAAFVASTLMEYGGVPQSATPESLLKEAIHVISCGYEDKIEWGTEIGWIYGSVTED 779

Qy      678 ILSGFKMHCRGWRSIYCMPVRPAFKGSAPINLSDRLHQVLRWALGSVEIFFSRHCPLWYG 737
            ||:|||:| ||||:|||| |||||||||||||||:|||||||||||| ||||||||||
Db      780 ILTGFKMHARGWRSIYCMPKRPAFKGSAPINLSDRLNQVLRWALGSVEILFSRHCPLWYG 839

Qy      738 FGGGRLKWLQRLAYINTIVYPFTSLPLIAYCTIPAVCLLTGKFIIPTLSNLASMLFLGLF 797
            : |||||:|:| |||||  :|| |||||: || :|;:|||||||||| :|| :|: |: |'
Db      840 Y-GGRLKFLERFAYINTTIYPLTSLPLLIYCILPAICLLTGKFIIPEISNFASIWFISLF 898

Qy      798 ISIIVTAVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQGFLKMLAGIDTNFTVTAK 857
            |||  | :||:|||||  |:: ||||||:||||:|||||||| ||:|||||||||:|
Db      899 ISIFATGILEMRWSGVGIDEWWRNEQFWVIGGISAHLFAVFQGLLKVLAGIDTNFTVTSK 958

Qy      858 AA-EDAEPGELYMVKWTTLLIPPTTLLIINMSG-CAGFSDALNKGYEAWGPLFGKVFFAF 915
            |: || :| |||| '|||||:|||:|||: :  |: | |:| ||::||||||:|||
Db      959 ASDEDGDFAELYMFKWTTLLIPPTTILIINLVGVVAGISYAINSGYQSWGPLFGKLFFAF 1018

Qy      916 WVILHLYPFLKGLMGRQNLTPTIVVLWSVLLASVFSLVWVKINPFVNKVDNTLVAETCIS 975
            |||:|||||||| '|| |,|| ::|::|| |:|||:||:|:|| |:|        ||
Db     1019 WVIVHLYPFLKGLMGRQNRTPTIVVVWAILLASIFSLLWVRIDPFTNRVTGP-DTRTC-G 1076

Qy      976 IDC 978
            |:|
Db     1077 INC 1079


RESULT 14
Q9XGX6
ID   Q9XGX6      PRELIMINARY;      PRT;  1067 AA.
AC   Q9XGX6;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Cellulose synthase catalytic subunit.
GN   CELA3.
OS   Gossypium hirsutum (Upland cotton).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
```

```
CC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium.
OX   NCBI_TaxID=3635;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Texas marker-1; TISSUE=Fiber;
RX   MEDLINE=20098737; PubMed=10631273;
RA   Laosinchai W., Cui X., Brown R.M. Jr.;
RT   "A full length cDNA of cotton cellulose synthase has high homology
RT   with the Arabidopsis RSW1 gene and the cotton CelA1 gene (Accession
RT   No. AF200453) (PGR 00-002).";
RL   Plant Physiol. 122:291-291(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Texas marker-1; TISSUE=Fiber;
RA   Kimura S., Laosinchai W., Itoh T., Cui X., Brown R.M. Jr.;
RT   "Immunogold Labeling of Rosette Terminal Cellulose Synthesizing
RT   Complexes in a Vascular Plant (Vigna angularis).";
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF150630; AAD39534.2; -.
DR   InterPro; IPR005150; Cellulose_synt.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF03552; Cellulose_synt; 1.
DR   SMART; SM00184; RING; 1.
SQ   SEQUENCE   1067 AA;  119325 MW;  0BBA2ED00590F29C CRC64;

  Query Match             64.2%;  Score 3366;  DB 10;  Length 1067;
  Best Local Similarity   59.8%;  Pred. No. 5.7e-257;
  Matches  645;  Conservative  133;  Mismatches  177;  Indels  124;  Gaps   20;

Qy          1 MMESGAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEINEGRKVCLRC------ 54
              |    |  |  ||: || : :|: |:|:: |:::|: :|:|   :| : | :|
Db         12 MKNLGGQTCQICGDNVGKNTDGDPFIACNICAFPVCRPCYEYERKDGNQSCPQCKTRYKW 71

Qy         55 --GSP----------YDENLLDDVEKKGSGNQSTMASHLNI  -  ----SQDVG------ 88
                 |||            |:   | :  :   :  :|  :           :|:|
Db         72 QKGSPAILGDRETGGDADDGASDFIYSENQEQKQKLAERMQ WNAKYGRGEDVGAPTYDK 131

Qy         89  IHARHI---------------  --- -  ---   SSVSTVDSEM---NDEYGNPI 132
               |   ||                            !!   !!     :   ||
Db        132 EISRNHIPLLTSGQEVSGELSAASPERLSMASPGVAGGKSSIRVVDPVREFGSSGLGNVA 191

Qy        113 WKNRVKSCKDKENKKKKRSPKAETEP   AQVPTEQQM--   --  --- EEKPSAEASE 157
              || ||   | |: |          | |   |  :|: :                : : : || :
Db        192 WKERVDGWKMKQEK  ------NTVPMSTCQATSERGIGDI GSTDVLVDDSQLNDEARQ 243

Qy        158 PLSIVYPIPRNKLTPYRAVIIMRLVILGLFFHFRITNPVA  HWLISVICEIWFAFSW 217
              |:'     :|: |||  :::|:|| :  :|   :||  |  :   ||  |||:.||  ||
Db        244 PLSREVSVSSSKINPYRMVIIIRLVILALFLHYRITNPVN AIWLISVICEIWFAISW 303

Qy        218 VLDQFPKWNPVNRETYIERLSARYEREGEPSQLAAVDIFVS VDPLKEPPLITANTVLSI 277
              :|:|||| |||||||::||: :'|||||':|| |: ||: |||:|||||:||||||||
Db        304 ILDQFPKWLPVNRETYLDRLALRYDREGEPSELAAVDIFVS VDPLKEPPLVTANTVLSI 363

Qy        278 LAVDYPVDKVSCYVSDDGAAMLSFESLVETAEFARKWVPFC KFSIEPRAPEFYFSQKID 337
              |'|||||||||||||||||||:||:| ||:|:|||||:|:|:::|||:|:|:||||
Db        364 LAVDYPVDKVSCYVSDDGAAMLTFEALSETSEFARKWVPFC KYNIEPRAPEWYFAQKID 423

Qy        338 YLKDKVQPSFVKERRAMKRDYEEYKVRVNALVAKAQKTPEE WTMQDGTPWPGNNTRDHP 397
              |||||||:||||||||:|:|:|:|:| |'|||| ||||||||||||||||||||||||
Db        424 YLKDKVQTSFVKDRRAMKREYEEFKVRINGLVAKAQKVPEE WIMQDGTPWPGNNTRDHP 483

Qy        398 -------GHDSGLPWEILGARDIEGNELPRLVYVSREKRPGY QHHKKAGAENALVRVSAV 450
                    |   ||         | ||||||||||||||||:| |:|||||  ||||||||
Db        484 GMIQVFLGQSGGL--------DAEGNELPRLVYVSPEKRPGF QHHKKAGAMNALVRVSAV 535

Qy        451 LTNAPYILNVDCDHYVNNSKAVREAMCILMDPQVGRDVCYV QPQRFDGIDKSDRYANRN 510
              ||| ::'|:|||| :|||||:||||| |||: :|: |||| ||||||||::|||||||
Db        536 LTNGAFLLNLDCDHYINNSKALREAMCFLMDPNLGKQVCYV QPQRFDGIDRNDRYANRN 595

Qy        511 VVFFDVNMKGLDGIQGPVYVGTGCVFNRQALYGYDPPSMPS KRKDSSSCFSCCCPSKK 570
              |||:|::|||||||||||| ||| :||| :|  '  ||    :   | |  |:'
Db        596 IVFFDINLRGLDGIQGPVYVGTGCVFNRTALYGYEPPLKPK  K   TGILSSLCGGSRK 652
```

```
Qy        571 KPAQDP---AEVYRDAKREDLNAAIFNLTEID------NYDEHERSMLISQLSFEKTFGL 621
              | ::     ::  :  |  |    :||| :|:       :|: |:|:|:||:| || ||
Db        653 KSSKSSKKGSDKKKSGKHVDSTVPVFNLEDIEEGVEGAGFDD-EKSLLMSQMSLEKRFGQ 711

Qy        622 SSVFIESTLMENGGVPESANSPPFIKEAIQVIGCGYEEKTEWGKQIGWIYGSVTEDILSG 681
              |:||: ||||||||||:||     :|||| || ||||:||:|| :|||||||||||:|
Db        712 SAVFVASTLMENGGVPQSATPETLLKEAIHVISCGYEDKTDWGSEIGWIYGSVTEDILTG 771

Qy        682 FKMHCRGWRSIYCMPVRPAFKGSAPINLSDRLHQVLRWALGSVEIFFSRHCPLWYGFGGG 741
              |||| |||||||||| |||||||||||||||:|||||||||||| ||||||:|||:  |
Db        772 FKMHARGWRSIYCMPKRPAFKGSAPINLSDRLNQVLRWALGSVEILFSRHCPIWYGY-SG 830

Qy        742 RLKWLQRLAYINTIVYPFTSLPLIAYCTIPAVCLLTGKFIIPTLSNLASMLFLGLFISII 801
              |||||:| ||:|| :|| |::||: |||:||||||| ||||| :|||||: |: ||:||
Db        831 RLKWLERFAYVNTTIYPVTAIPLLMYCTLPAVCLLTNKFIIPQISNLASIWFISLFLSIF 890

Qy        802 VTAVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQGFLKMLAGIDTNFTVTAKAA-E 860
              | :|:::|:|| |:  |||||||||||||||||||||| ||:|||||||||:||: |
Db        891 ATGILKMKWNGVGIDQWWRNEQFWVIGGVSAHLFAVFQGLLKVLAGIDTNFTVTSKASDE 950

Qy        861 DAEFGELYMVKWTTLLIPPTTLLIINMSG-CAGFSDALNKGYEAWGPLFGKVFFAFWVIL 919
              | :| |||| ||||||||||||||||: |  || |  :| ||::|||||||:|||||||:
Db        951 DGDFAELYMFKWTTLLIPPTTLLIINLVGVVAGISYVINSGYQSWGPLFGKLFFAFWVII 1010

Qy        920 HLYPFLKGLMGRQNLTPTIVVLWSVLLASVFSLVWVKINPFVNKVDNTLVAETCISIDC 978
              |||||||||||||| ||||||:||:||||:|||:||:|:||  :|    | | |  |:|
Db       1011 HLYPFLKGLMGRQNRTPTIVVVWSILLASIFSLLWVRIDPFTTRVTGPDV-EQC-GINC 1067
```

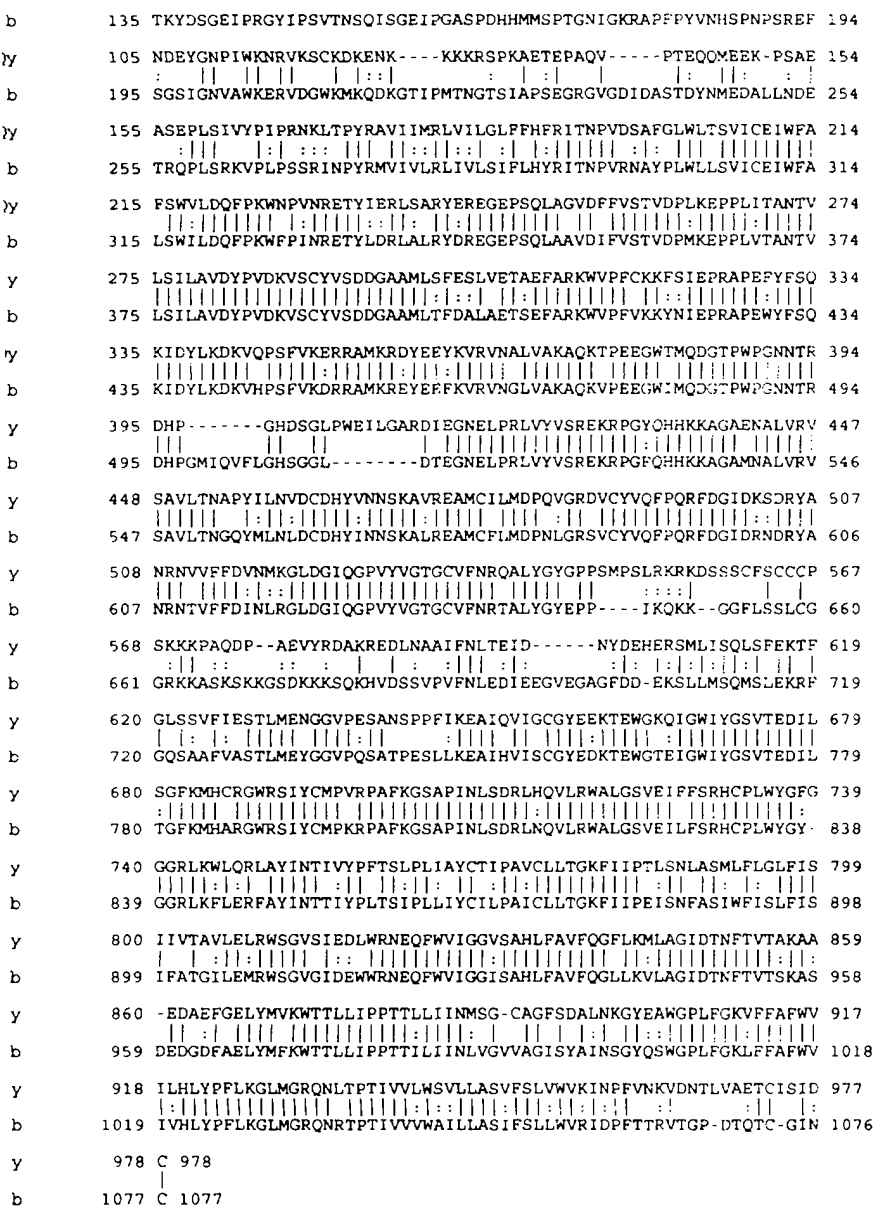
```
RESULT 15
Q9LLI6
ID   Q9LLI6      PRELIMINARY;      PRT;  1077 AA.
AC   Q9LLI6;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Cellulose synthase-4.
GN   CESA-4.
OS   Zea mays (Maize).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   PACCAD clade; Panicoideae; Andropogoneae; Zea.
OX   NCBI_TaxID=4577;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20398328; PubMed=10938350;
RA   Holland N., Holland D., Helentjaris T., Dhugga K.S.,
RA   Xoconostle-Cazares B., Delmer D.P.;
RT   "A comparative analysis of the plant cellulose synthase (CesA) gene
RT   family.";
RL   Plant Physiol. 123:1313-1324(2000).
DR   EMBL; AF200528; AAF89964.1; -.
DR   InterPro; IPR005150; Cellulose_synt.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF03552; Cellulose_synt; 1.
DR   SMART; SM00184; RING; 1.
SQ   SEQUENCE   1077 AA;  120501 MW;  5F960E4AA753E2D6 CRC64;

  Query Match             64.2%;  Score 3365.5;  DB 10;  Length 1077;
  Best Local Similarity   59.7%;  Pred. No. 6.3e-257;
  Matches  645;  Conservative  136;  Mismatches  175;  Indels  125;  Gaps   20;

Qy          5 GAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEINEGRKVCLRC--------GS 56
              |  :|  ||: ||  | |::| ||  | :|:|: |:|:|  :| : | :|         ||
Db         15 GGQVCQICGDGVGTTAEGDVFAACDVCGFPVCRPCYEYERKDGTQACPQCKTKYKRHKGS 74

Qy         57 P---  ----------YDENLL----DDVEKKGSGNQSTMASHLNDSQDVG---------- 88
                  |              | | |    :| ::| :    :   ::  | |||
Db         75 PAIRGEEGDDTDADSDFNYLASGNEDQKQKIADRMRSWRMNVGGSGDVGRPKYDSGEIGL 134

Qy         89 --  -  IHARHISSVS------------ ----------------------TVDSEM 104
                    |   :| ||:                                      |
```



earch completed: October 29, 2003, 18:30:50
ob time : 78.1006 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 31, 2003, 09:33:04 ; Search time 5802.74 Seconds
(without alignments)
4096.305 Million cell updates/sec

Title: US-09-980-043-2
Perfect score: 5240
Sequence: 1 MMESGAPICHTCGEQVGHDA.....PFVNKVDNTLVAETCISIDC 978

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09980043/runat_29102003_163853_21871/app_query.fasta_1.1358
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL 1 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX 100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09980043_@CGN_1_1_4125_@runat_29102003_163853_21871 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3365.5	64.2	3763	11	AY108113	AY108113 Zea mays
2	3361	64.1	3783	11	AY104730	AY104730 Zea mays
3	3333.5	63.6	3897	11	AY103655	AY103655 Zea mays
4	3304.5	63.1	3788	11	AY103701	AY103701 Zea mays
5	3258.5	62.2	3898	11	AY110415	AY110415 Zea mays
6	3149	60.1	2872	11	AY104236	AY104236 Zea mays
7	3105.5	59.3	3728	11	AY112236	AY112236 Zea mays
8	3067.5	58.5	3696	11	AY110079	AY110079 Zea mays
9	1378	26.3	1189	11	AY107656	AY107656 Zea mays
10	1263.5	24.1	953	14	CD439980	CD439980 EL01N0531
11	1231	23.5	960	9	AB049602	AB049602 AB049602
12	1219	23.3	1022	29	BZ827769	BZ827769 PUFHO17TB
13	1197.5	22.9	792	14	CB671068	CB671068 OSJNEe04H
14	1188.5	22.7	851	14	CB892280	CB892280 EST649249
15	1187	22.7	731	13	BQ867361	BQ867361 QGD10J04.
16	1180.5	22.5	799	14	CB894714	CB894714 EST647506
17	1177	22.5	735	13	BQ969585	BQ969585 QHB38P07.
18	1169	22.3	800	14	CB648832	CB648832 OSJNEb12G
19	1144.5	21.8	753	14	CA784277	CA784277 sat96g06.
20	1120.5	21.4	779	14	CA916873	CA916873 EST641020
21	1110.5	21.2	807	12	BI311236	BI311236 EST531298
22	1101.5	21.0	788	10	BG584483	BG584483 EST486243
23	1098.5	21.0	688	13	BU099300	BU099300 WHE3305_H
24	1087	20.7	862	14	CD437450	CD437450 EL01N0501
25	1086	20.7	978	10	BG321258	BG321258 Zm04_03f1
26	1084.5	20.7	807	14	CA918393	CA918393 EST642540
c 27	1084	20.7	984	29	BZ792792	BZ792792 PUGDK50TB
28	1080.5	20.6	783	14	CB671503	CB671503 OSJNEe05B
29	1080	20.6	718	13	BU896705	BU896705 X044E08 P
30	1074	20.5	830	14	CA917698	CA917698 EST641845
31	1072.5	20.5	856	14	CB633016	CB633016 OSIIEb11J
32	1068.5	20.4	731	13	BU007898	BU007898 QGH5M24.y
c 33	1060.5	20.2	725	13	BQ965643	BQ965643 QHB22I09.
34	1052.5	20.1	717	13	BQ996712	BQ996712 QGG13J11.
35	1050.5	20.0	668	10	BF635224	BF635224 NF061H04D
36	1049.5	20.0	714	14	CB629928	CB629928 OSIIEb06I
37	1045.5	20.0	741	10	BF460078	BF460078 072G01 Ma
38	1041.5	19.9	712	13	BU004334	BU004334 QGG4e11.y
39	1040.5	19.9	792	12	BI925066	BI925066 EST544955
40	1039.5	19.8	880	12	BM816138	BM816138 HC109B12_
41	1033	19.7	787	14	CD486470	CD486470 CRH2.4E12
42	1032.5	19.7	685	13	BX249614	BX249614 BX249614
43	1025.5	19.6	712	13	BQ620630	BQ620630 TaLr1138D
44	1023	19.5	715	13	BQ870049	BQ870049 QGD7P10.y
45	1023	19.5	842	12	BI311476	BI311476 EST531322

ALIGNMENTS

RESULT 1
AY108113
LOCUS AY108113 3763 bp mRNA linear HTC 16-OCT-2002
DEFINITION Zea mays PCO126465 mRNA sequence.
ACCESSION AY108113
VERSION AY108113.1 GI:21211191
KEYWORDS HTC.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 3763)

UTHORS Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S., Arthur,L.W., Hanafey,M., Morgante,M and Tingey,S.V.
ITLE Maize Mapping Project/DuPont Consensus Sequences for Design of Overgo Probes
OURNAL Unpublished (2002)
ERENCE 2 (bases 1 to 3763)
UTHORS Coe,E.H.
ITLE Direct Submission
OURNAL Submitted (25-APR-2002) Maize Mapping Project, University of Missouri, Columbia, MO 65211, USA
MENT If you are interested in getting corresponding physical clones, these are publicly available from ZmDB and may be found by BLAST searching at MSL, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR, www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the maize cDNA sequences is either Virginia Walbot, Stanford or Pat Schnable, Iowa State, then clones may be requested from ZmDB: www.zmdb.iastate.edu.

```
ATURES             Location/Qualifiers
   source          1. .3763
                   /organism="Zea mays"
                   /mol_type="mRNA"
                   /db_xref="MaizeDB:637367"
                   /db_xref="taxon:4577"
                   /clone_lib="Maize Mapping Project/DuPont Cornsensus
                   Library"
                   /note="this sequence is part of a project of EST
                   assemblies resulting from the application of public
                   contigs to seed DuPont contigs; this resource was
                   assembled by DuPont as part of a collaboration for the
                   overgo addressing of BACs in conjunction with the Maize
                   Mapping Project"
SE COUNT      863 a    984 c   1011 g    904 t      1 others
IGIN
```

ignment Scores:

ed. No.:	1.31e-286	Length:	3763
ore:	3365.50	Matches:	648
rcent Similarity:	73.00%	Conservative:	136
st Local Similarity:	60.34%	Mismatches:	180
ery Match:	64.23%	Indels:	112
:	11	Gaps:	18

3-09-980-043-2 (1-978) x AY108113 (1-3763)

```
/      5 GlyAlaProIleCysHisThrCysGlyGluGlnValGlyHisAspAlaAsnGlyGluLeu 24
       380 GGCGGACAGGTGTGCCAGATCTGCGGCGACGGCGTGGGCACCACGGCGGAGGGGGACGTC 439

y      25 PheValAlaCysHisGluCysSerTyrProMetCysLysSerCysPheGluPheGluIle 44
b      440 TTCGCCGCCTGCGACGTCTGCGGGTTTCCGGTGTGCCGCCCCTGCTACGAGTACGAGCGC 499

y      45 AsnGluGlyArgLysValCysLeuArgCys----------------------GlySer 56
b      500 AAGGACGGCACGCAGGCGTGCCCCCAGTGCAAGACCAAGTACAAGCGCCACAAGGGGAGC 559

y      57 Pro-----------------------------------------TyrAspGluAsnLeuLeu 63
b      560 CCGGCGATCCGTGGGGAGGAAGGAGACGACACTGATGCCGATAGCGACTTCAATTACCTT 619

Jy     64 ------------AspAspValGluLysLysGlySerGlyAsnGlnSerThrMetAlaSer 79
Jb     620 GCATCTGGCAATGAGGACCAGAAGCAGAAGATTGCCGACAGAATGCGCAGCTGGCGCATG 679

Jy     80 HisLeuAsnAspSerGlnAspValGly------------------------------- 88
Jb     680 AACGTTGGGGGCAGCGGGGATGTTGGTCGCCCCAAGTATGACAGTGGCGAGATCGGGCTT 739

Jy     89 --------------------IleHisAlaArgHisIleSerSerValSer--------- 98
Jb     740 ACCAAGTATGACAGTGGCGAGATTCCTCGGGGATACATCCCATCAGTCACTAACAGCCAG 799

Jy     98 ------------------------------------------------------------ 98
```

```
Db     800 ATCTCAGGAGAAATTCCTGGTGCTTCCCCTGACCATCATATGATGTCCCCAACTGGGAAC 859

Qy      99 ------------ ------ - -- -- -- ----------ThrValAspSerGluMet 104

Db     860 ATTGGCAAGCGTGCTCCATTTCCCTATGTGAACCATTCGCCAAATCCGTCAAGGGAGTTC 919

Qy     105 AsnAspGluTyrGlyAsnProIleTrpLysAsnArgValLysSerCysLysAspLysGlu 124

Db     920 TCTGGTAGCATTGTAATGTTGCCTGGAAAGAGAGGGTTGATGGCTGGAAAATGAAGCAG 979

Qy     125 AsnLys--- --- -- LysLysLysArgSerProLysAlaGluThrGluProAlaGln 140

Db     980 GACAAGGGGACGATCCCATGACGAATGGCACAAGCATTGCTCCCTCTGAGGGTCGGGGT 1039

Qy     141 Val---------- -ProThrGluGlnGlnMetGluGluLys---ProSerAlaGlu 154

Db     1040 GTTGGTGATATTGACGATCAACTGATTACAACATGGAAGATGCCTTATTGAACGACGAA 1099

Qy     155 AlaSerGluProI Ser IleValTyrProIleProArgAsnLysLeuThrProTyrArg 174

Db     1100 ACTCGACAGCTCTCTATAGAAAATTCCACTCCTCCTTCCTCCAGGATAAATCCATACAGG 1159

Qy     175 AlaValIleIleM ArgLeuValIleLeuGlyLeuPhePheHisPheArgIleThrAsn 194

Db     1160 ATGGTCATTGTGCTCGATTGATTGTTCTAAGCATCTTCTTGCACTACCGTATCACAAAT 1219

Qy     195 ProValAspSerAlaPheGlyLeuTrpLeuThrSerValIleCysGluIleTrpPheAla 214

Db     1220 CCTGTGCCGCAATGCATACCCATTATGGCTTCTATCTGTTATATGTGAGATCTGGTTTGCT 1279

Qy     215 PheSerTrpValLeuAspGlnPheProLysTrpAsnProValAsnArgGluThrTyrIle 234

Db     1280 CTTTCGTGGATATTCGATCAGTTCCCTAAGTGGTTTCCAATCAACCGGGAGACGTACCTT 1339

Qy     235 GluArgLeuSerAlaArgTyrGluArgGluGlyGluProSerGlnLeuAlaGlyValAsp 254

Db     1340 GATAGGCTGGCATTAAGGTATGACCGGGAAGGTGAGCCATCTCAGTTGGCTGCTGTTGAC 1399

Qy     255 PhePheValSerThrValAspProLeuLysGluProProLeuIleThrAlaAsnThrVal 274

Db     1400 ATTTTCGTCAGTACAGTCGACCCAATGAAGGAGCCTCCTCTTGTCACTGCCAATACCGTG 1459

Qy     275 LeuSerIleLeuAlaValAspTyrProValAspLysValSerCysTyrValSerAspAsp 294

Db     1460 CTATCCATTCTTGCTGTGGATTACCCTGTGGATAAGGTCTCTTGCTATGTATCTGATGAT 1519

Qy     295 GlyAlaAlaMetLeuSerPheGluSerLeuValGluThrAlaGluPheAlaArgLysTrp 314

Db     1520 GGAGCTGCGATGCTGACATTTGATGCACTAGCTGAGACTTCAGAGTTTGCTAGAAAATGG 1579

Qy     315 ValProPheCysLysLysPheSerIleGluProArgAlaProGluPheTyrPheSerGln 334

Db     1580 GTACCATTTGTTAAGAAGTACAACATTGAACCTAGAGCTCCTGAATGGTACTTCTCCCAG 1639

Qy     335 LysIleAspTyrLeuLysAspLysValGlnProSerPheValLysGluArgArgAlaMet 354

Db     1640 AAAATTGATTACTTGAAGGACAAAGTGCACCCTTCATTTGTTAAAGACCGCCGGGCCATG 1699

Qy     355 LysArgAspTyrGlnGluTyrLysValArgValAsnAlaLeuValAlaLysAlaGlnLys 374

Db     1700 AAGAGAGAATATGAAGAATTCAAAGTTAGGGTAAATGGCCTTGTTGCTAAGGCACAGAAA 1759

Qy     375 ThrProGluGluGlyTrpThrMetGlnAspGlyThrProTrpProGlyAsnAsnThrArg 394

Db     1760 GTTCCTGAGGAAGGATGGATCATGCAAGATGGCACACCATGGCCAGGAAACAATACCMGG 1819

Qy     395 AspHisProGlyHisAspSerGlyLeuProTrpGluIleLeuGlyAlaArgAspIleGlu 414

Db     1820 GACCATCCTGGAAT-GATTCAGGTTTTCCTTGGTCAC--AGTGGTGGCCTTGATACTGAG 1876

Qy     415 GlyAsnGluLeuProArgLeuValTyrValSerArgGluLysArgProGlyTyrGlnHis 434
```

```
Db       1877 GGCAATGAGCTACCCCGTTTGGTCTATGTTTCTCGTGAAAAGCGTCCTGGATTCCAGCAT 1936
Qy        435 HisLysLysAlaGlyAlaGluAsnAlaLeuValArgValSerAlaValLeuThrAsnAla 454
Db       1937 CACAAGAAAGCTGGTGCCATGAATGCTCTTGTTCGTGTCTCAGCTGTGCTTACCAATGGA 1996
Qy        455 ProTyrIleLeuAsnValAspCysAspHisTyrValAsnAsnSerLysAlaValArgGlu 474
Db       1997 CAATACATGTTGAATCTTGATTGTGATCACTACATTAACAACAGTAAGGCTCTCAGGGAA 2056
Qy        475 AlaMetCysIleLeuMetAspProGlnValGlyArgAspValCysTyrValGlnPhePro 494
Db       2057 GCTATGTGCTTCCTTATGGACCCTAACCTAGGAAGGAGTGTCTGCTACGTCCAGTTTCCC 2116
Qy        495 GlnArgPheAspGlyIleAspLysSerAspArgTyrAlaAsnArgAsnValValPhePhe 514
Db       2117 CAGAGATTCGATGGCATTGACAGGAATGATCGATATGCCAACAGGAACACCGTGTTTTTC 2176
Qy        515 AspValAsnMetLysGlyLeuAspGlyIleGlnGlyProValTyrValGlyThrGlyCys 534
Db       2177 GATATTAACTTGAGAGGTCTTGATGGCATCCAAGGACCAGTTTATGTCGGAACTGGCTGT 2236
Qy        535 ValPheAsnArgGlnAlaLeuTyrGlyTyrGlyProProSerMetProSerLeuArgLys 554
Db       2237 GTTTTCAACCGAACAGCTCTATATGGTTATGAGCCCCCA-- ---------ATTAAGCAG 2284
Qy        555 ArgLysAspSerSerSerCysPheSerCysCysCysProSerLysLysLysProAlaGln 574
Db       2285 AAGAAG------GGTGGTTTCTTGTCATCACTATGTGGCGGTAGGAAGAAGGCAAGCAAA 2338
Qy        575 AspPro------AlaGluValTyrArgAspAlaLysArgGluAspLeuAsnAlaAlaIle 592
Db       2339 TCAAAGAAGGGCTCGGACAAGAAGAAGTCGCAGAAGCATGTGGACAGTTCTGTGCCAGTA 2398
Qy        593 PheAsnLeuThrGluIleAsp---------- ---AsnTyrAspGluHisGluArg 606
Db       2399 TTCAACCTTGAAGATATAGAGGAGGGAGTTGAAGGCGCTGGATTTGACGAC---GAGAAA 2455
Qy        607 SerMetLeuIleSerGlnLeuSerPheGluLysThrPheGlyLeuSerSerValPheIle 626
Db       2456 TCACTTCTTATGTCTCAAATGAGCCTGGAGAAGAGATTTGGCCAGTCCGCAGCGTTTGTT 2515
Qy        627 GluSerThrLeuMetGluAsnGlyGlyValProGluSerAlaAsnSerProProPheIle 646
Db       2516 GCCTCCACTCTGATGGAGTATGGTGGTGTTCCTCAGTCCGCAACTCCGGAGTCTCTTCTG 2575
Qy        647 LysGluAlaIleGlnValIleGlyCysGlyTyrGluGluLysThrGluTrpGlyLysGln 666
Db       2576 AAAGAAGCTATCCATGTTATAAGCTGTGGCTATGAGGAAAA ACTGAATGGGGAACTGAG 2635
Qy        667 IleGlyTrpIleTyrGlySerValThrGluAspIleLeuSerGlyPheLysMetHisCys 686
Db       2636 ATCGGGTGGATCTACGGTTCTGTGACAGAAGACATTCTCACCGGATTCAAGATGCACGCG 2695
Qy        687 ArgGlyTrpArgSerIleTyrCysMetProValArgProAlaPheLysGlySerAlaPro 706
Db       2696 CGAGGCTGGCGGTCGATCTACTGCATGCCCAAGCGGCCAGCTTTCAAGGGGTCTGCCCCC 2755
Qy        707 IleAsnLeuSerAspArgLeuHisGlnValLeuArgTrpAlaLeuGlySerValGluIle 726
Db       2756 ATCAATCTTTCGGACCGTCTGAACCAGGTGCTCCGGTGGGCCCTTGGGTCCGTGGAGATC 2815
Qy        727 PhePheSerArgHisCysProLeuTrpTyrGlyPheGlyGlyArgLeuLysTrpLeu 746
Db       2816 CTCTTCAGCCGGCACTGCCCCCTGTGGTACGGCTAC---GGAGGGCGGCTCAAGTTCCTG 2872
Qy        747 GlnArgLeuAlaTyrIleAsnThrIleValTyrProPheThrSerLeuProLeuIleAla 766
Db       2873 GAGAGATTCGCGTACATCAACACCACCATCTACCCGCTCACGTCCATCCCGCTTCTCATC 2932
Qy        767 TyrCysThrIleProAlaValCysLeuLeuThrGlyLysPheIleIleProThrLeuSer 786
Db       2933 TACTGCATCCTGCCCGCCATCTGTCTGCTCACCGGAAAGTTCATCATTCCAGAGATCAGC 2992

Qy        787 AsnLeuAlaSerMetLeuPheLeuGlyLeuPheIleSerIleIleValThrAlaValLeu 806
Db       2993 AACTTCGCCAGCATCTGGTTCATCTCCCTCTTCATCTCGATCTTCGCCACGGGCATCCTG 3052
Qy        807 GluLeuArgTrpSerGlyValSerIleGluAspLeuTrpArgAsnGluGlnPheTrpVal 826
Db       3053 GAGATGAGGTGGAGCGGGGTGGGCATCGACGAGTGGTGGAGGAACGAGCAGTTCTGGGTG 3112
Qy        827 IleGlyGlyValSerAlaHisLeuPheAlaValPheGlnGlyPheLeuLysMetLeuAla 846
Db       3113 ATCGGGGGCATCTCCGCGCACCTCTTCGCCGTGTTCCAGGGCCTGCTCAAGGTGCTGGCC 3172
Qy        847 GlyIleAspThrAsnPheThrValThrAlaLysAlaAla---GluAspAlaGluPheGly 865
Db       3173 GGCATCGACACCAACTTCACCGTCACCTCCAAGGCCTCGGACGAGGACGGCGACTTCGCG 3232
Qy        866 GluLeuTyrMetValLysTrpThrThrLeuLeuIleProProThrThrLeuLeuIleIle 885
Db       3233 GAGCTGTACATGTTCAAGTGGACGACGCTCCTGATCCCGCCCACCACCATCCTGATCATC 3292
Qy        886 AsnMetSerGly---CysAlaGlyPheSerAspAlaLeuAsnLysGlyTyrGluAlaTrp 904
Db       3293 AACCTGGTCGGCGTCGTCGCCGGCATCTCCTACGCCATCAACAGCGGATACCAGTCGTGG 3352
Qy        905 GlyProLeuPheGlyLysValPhePheAlaPheTrpValIleLeuHisLeuTyrProPhe 924
Db       3353 GGCCCGCTCTTCGGCAAGCTCTTCTTCGCCTTCTGGGTCATCGTCCACCTGTACCCGTTC 3412
Qy        925 LeuLysGlyLeuMetGlyArgGlnAsnLeuThrProThrIleValValLeuTrpSerVal 944
Db       3413 CTCAAGGGCCTCATGGGCAGGCAGAACCGCACCCCGACCATCGTCGTCGTCTGGGCCATC 3472
Qy        945 LeuLeuAlaSerValPheSerLeuValTrpValLysIleAsnProPheValAsnLysVal 964
Db       3473 CTGCTGGCGTCCATCTTCTCCTTGCTGTGGGTTCGCATCGACCCCTTCACCACCCGCGTC 3532
Qy        965 AspAsnThrLeuValAlaGluThrCysIleSerIleAspCys 978
Db       3533 ACTGGCCCG---GATACCCAGACGTGT---GGCATCAACTGC 3568
```

RESULT 2
AY104730

```
LOCUS       AY104730               3783 bp    mRNA    linear   HTC 16-OCT-2002
DEFINITION  Zea mays PCO100501 mRNA sequence.
ACCESSION   AY104730
VERSION     AY104730.1  GI:21207808
KEYWORDS    HTC.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 3783)
  AUTHORS   Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S.,
            Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
  TITLE     Maize Mapping Project/DuPont Consensus Sequences for Design of
            Overgo Probes
  JOURNAL   Unpublished (2002)
REFERENCE   2  (bases 1 to 3783)
  AUTHORS   Coe,E.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-APR-2002) Maize Mapping Project, University of
            Missouri, Columbia, MO 65211, USA
COMMENT     If you are interested in getting corresponding physical clones,
            these are publicly available from ZmDB and may be found by BLAST
            searching at MSL, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR,
            www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the
            maize cDNA sequences is either Virginia Walbot, Stanford or Pat
            Schnable, Iowa State, then clones may be requested from ZmDB:
            www.zmdb.iastate.edu.
FEATURES             Location/Qualifiers
     source          1..3783
```

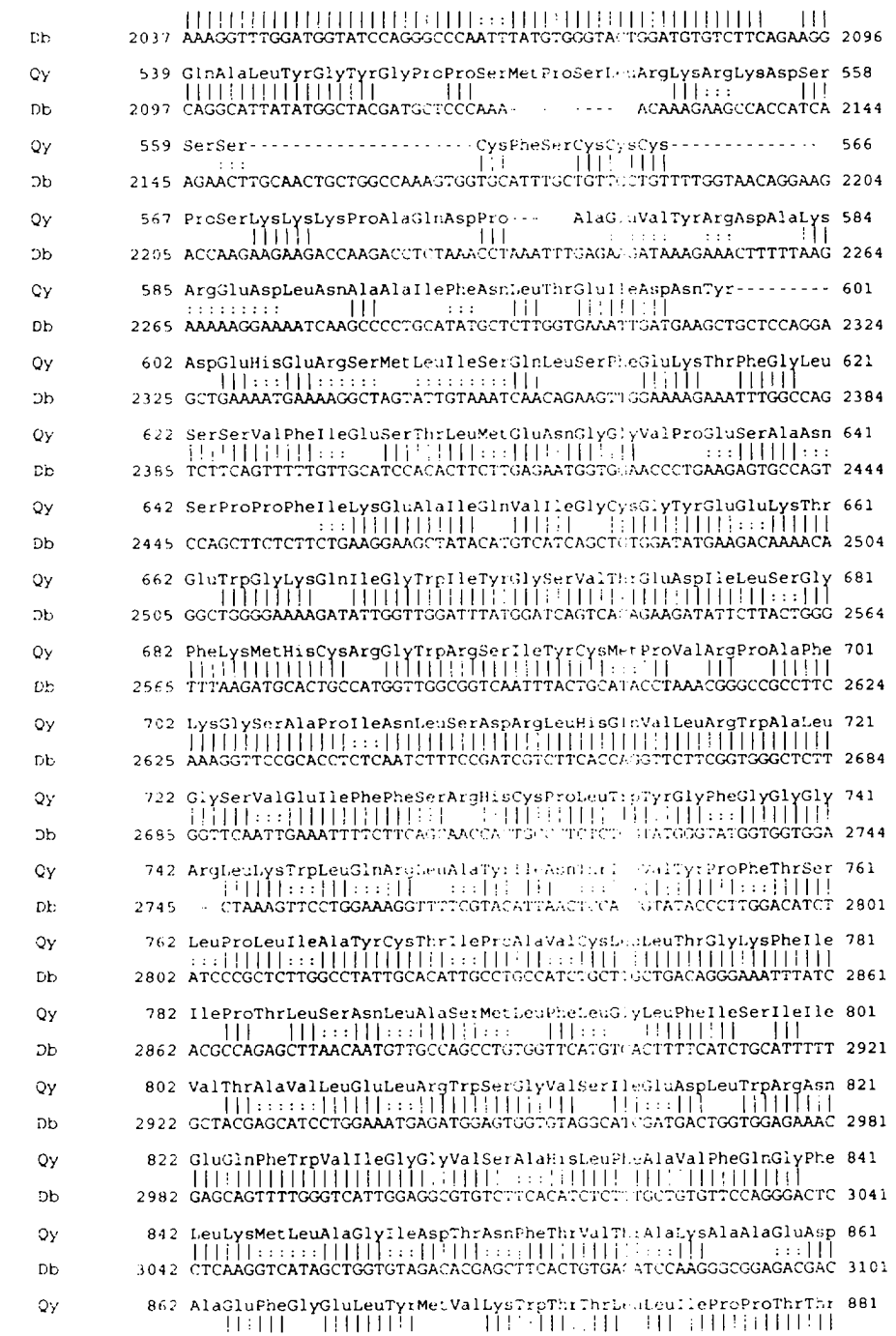
```
                /organism="Zea mays"
                /mol_type="mRNA"
                /db_xref="MaizeDB:635765"
                /db_xref="taxon:4577"
                /clone_lib="Maize Mapping Project/DuPont Cornsensus
                Library"
                /note="this sequence is part of a project of EST
                assemblies resulting from the application of public
                contigs to seed DuPont contigs; this resource was
                assembled by DuPont as part of a collaboration for the
                overgo addressing of BACs in conjunction with the Maize
                Mapping Project"
ASE COUNT      933 a    908 c   1002 g    940 t
RIGIN

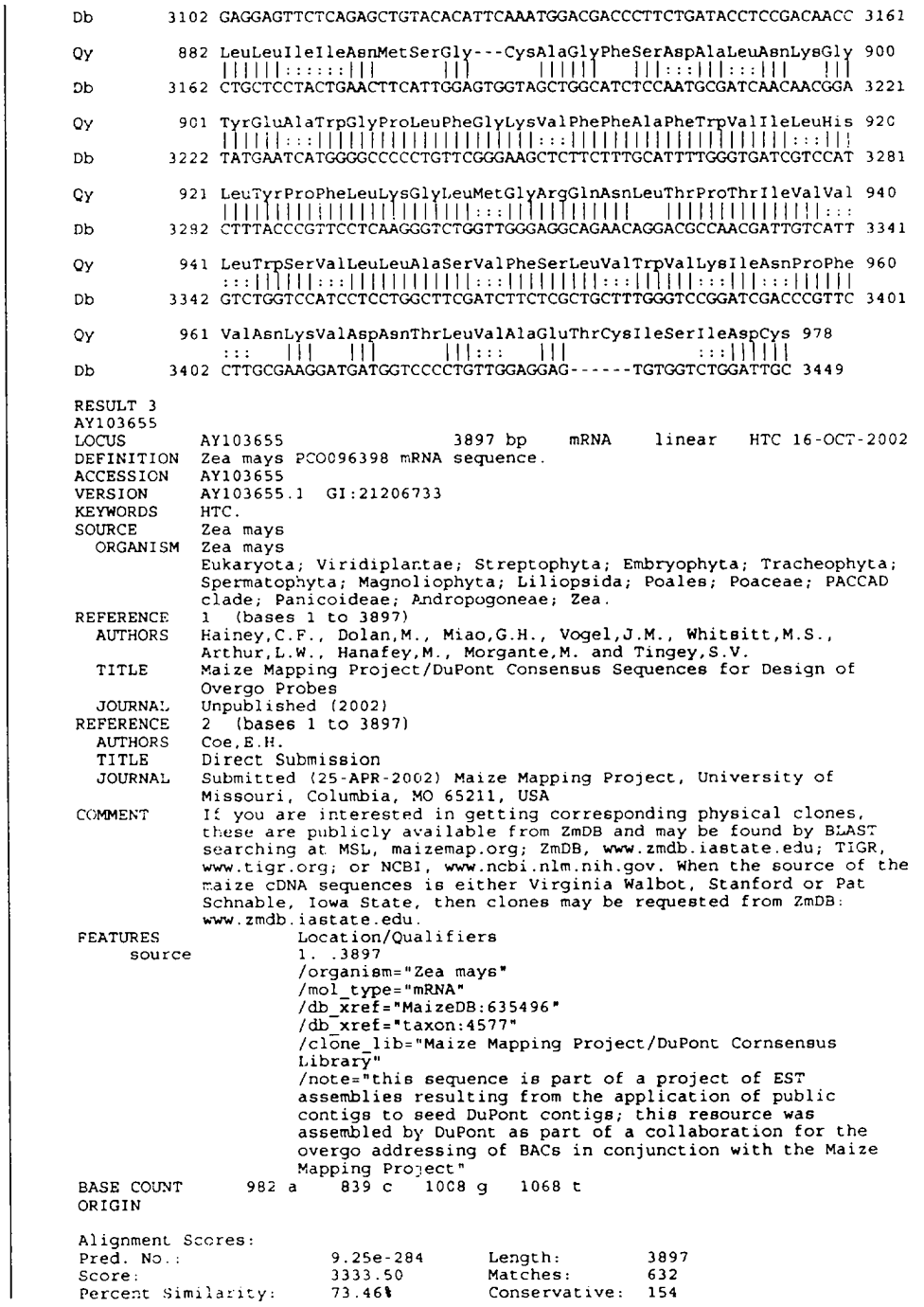
lignment Scores:
red. No.:                3.29e-286       Length:         3783
core:                    3361.00         Matches:        632
ercent Similarity:       72.63%          Conservative:   151
est Local Similarity:    58.63%          Mismatches:     176
uery Match:              64.14%          Indels:         121
:                        11              Gaps:           15

S-09-980 043-2 (1-978) x AY104730 (1-3783)

y          1 MetMetGluSerGlyAlaProIleCysHisThrCysGlyGluGlnValGlyHisAspAla 20
b        276 ATGGACCAGCGGAACGGCCAGGTGTGCCAGATTTGCGGCGACGACGTGGGGCGCAACCCC 335
y         21 AsnGlyGluLeuPheValAlaCysHisGluCysSerTyrProMetCysLysSerCysPhe 40
b        336 GACGGGGAGCCGTTCGTGGCCTGCAACGAGTGCGCCTTCCCCATCTGCCGGGACTGCTAC 395
y         41 GluPheGluIleAsnGluGlyArgLysValCysLeuArgCysGlySerProTyr------ 58
b        396 GAGTACGAGCGCCGCGAGGGCACGCAGAACTGCCCCCAGTGCAAGACCCGCTTCAAGCGC 455
y         59 --------------------------------AspGluAsnLeuLeuAspAspValGlu 67
b        456 CTCAAGGGGTGCGCGCGCGTGCCCGGGGACGAGGAGGAGGACGGCGTCGACGACCTGGAG 515
y         68 LysLysGlySerGlyAsnGlnSerThrMetAlaSerHisLeuAsnAspSerGlnAspVal 87
b        516 AACGAGTTCAACTGGAGCGACAAGCACGACTCCCAGTACCTCGCCGAGTCCATG------ 569
y         88 GlyIleHisAlaArgHisIleSerSerValSerThrValAsp---------------- 101
b        570 ---CTCCACGCC---CACATGAGCTACGGCCGCGGCGCCGACCTCGACGGCGTGCCGCAG 623
y        102 ------------------------------------SerGluMetAsnAspGlu--- 107
b        624 CCATTCCACCCCATCCCCAATGTTCCCCTCCTCACCAACGGACAGATGGTCGATGACATC 683
y        107 ------------------------------------------------------------ 107
b        684 CCGCCGGACCAGCACGCCCTTGTGCCCTCGTTCGTGGGTGGCGGGGGGAAGAGGATTCAC 743
y        107 ------------------------------------------------------------ 107
b        744 CCTCTCCCGTACGCGGATCCCAACCTTCCTGTGCAACCGAGGTCTATGGACCCTTCCAAG 803
y        108 -----------------TyrGlyAsnProIleTrpLysAsnArgValLysSerCysLys 121
b        804 GATCTCGCCGCATATGGCTACGGGAGCGTAGCATGGAAGGAGAGGATGGAGAGCTGGAAG 863
y        122 AspLysGluAsnLysLysLysLysArgSerProLysAla---------GluThrGluPro 138
b        864 CAGAAGCAGGAGAGGATGCACCAGACGAGGAACGATGGCGGCGGCGATGATGGTGATGAT 923
y        139 AlaGlnValProThrGluGlnGlnMetGluGluLysProSerAlaGluAlaSerGluPro 158
b        924 GCAGATCTACCACTAATGGAT----------------------GAAGCTAGACAGCCA 959

Qy       159 LeuSerIleValTyrProIleProArgAsnLysLeuThrProTyrArgAlaValIleIle 178
Db       960 TTGTCCAGAAAGAT CCGCTTCCTTCAAGCCAAATCAACCCCTATAGGATGATTATAATA 1019
Qy       179 MetArgLeuVal LeuGlyLeuPheHisPheArgIleThrAsnProValAspSer 198
Db      1020 ATTCGGCTAGTGG TTGTGTTTCTTCTTCCACTACCGAGTGATGCATCCGGTGCCTGAT 1079
Qy       199 AlaPheGlyLeuTrpLeuThrSerValIleCysGluIleTrpPheAlaPheSerTrpVal 218
Db      1080 GCATTTGCTTTATG GCTCATATCTGTGATCTGTGAAATTTGGTTTGCCATGTCTTGGATT 1139
Qy       219 LeuAspGlnPheP LysTrpAsnProValAsnArgGluThrTyrIleGluArgLeuSer 238
Db      1140 CTTGACCAGTTTCC AAAGTGGTTCCCTATCGAGAGGGAAACCTATCTTGACCGGCTGAGT 1199
Qy       239 AlaArgTyrGluAr GluGlyGluProSerGlnLeuAlaGlyValAspPhePheValSer 258
Db      1200 TTAAGGTTTGACAA GAAGGGCATCCTTCTCAACTCGCCCCTGTCGATTTCTTTGTCAGT 1259
Qy       259 ThrValAspProLe LysGluProProLeuIleThrAlaAsnThrValLeuSerIleLeu 278
Db      1260 ACGGTTGATCCCTT AAGGAACCTCCATTGGTCACTGCTAATACTGTTCTATCTATCCTT 1319
Qy       279 AlaValAspTyrPr ValAspLysValSerCysTyrValSerAspAspGlyAlaAlaMet 298
Db      1320 TCGGTGGATTATC AGTTGATAAGGTTTCATGCTACGTTTCTGATGATGGTGCTGCCATG 1379
Qy       299 LeuSerPheGluSerLeuValGluThrAlaGluPheAlaArgLysTrpValProPheCys 318
Db      1380 CTGACATTTGAAGCATTGTCTGAAACATCTGAATTTGCAAAGAAATGGGTTCCTTTCTGC 1439
Qy       319 LysLysPheSer IleGluProArgAlaProGluPheTyrPheSerGlnLysIleAspTyr 338
Db      1440 AAAAGATATAGCCT TGAGCCTCGTGCTCCAGAGTGGTACTTCCAACAGAAGATAGACTAC 1499
Qy       339 LeuLysAspLysValGlnProSerPheValLysGluArgArgAlaMetLysArgAspTyr 358
Db      1500 CTGAAAGACAAGGTTGGCGCCAAACTTTGTTAGAGAACGGAGAGCAATGAAGAGAGATAT 1559
Qy       359 GluGluTyrLysValArgValAsnAlaLeuValAlaLysAlaGlnLysThrProGluGlu 378
Db      1560 GAGGAATTCAAGGT AGAATCAATGCCTTGGTTGCTAAAGCCCAAAAGGTTCCTGAGGAA 1619
Qy       379 GlyTrpThrMetGlnAspGlyThrProTrpProGlyAsnAsnThrArgAspHisProGly 398
Db      1620 GGATGGACAATGCAGGATGGAACTCCATGGCCCGGAAATAATGTCCGTGATCATCCTGGA 1679
Qy       399 HisAspSerGlyLeuProTrpGluIleLeuGlyAlaArgAspIleGluGlyAsnGluLeu 418
Db      1680 AT-GATTCAGGTT TCCTTGGTCAA AGTGGTGGCCATGATGTGGAAGGAAATGAGCTG 1736
Qy       419 ProArgLeuValTyrValSerArgGluLysArgProGlyTyrGlnHisHisLysLysAla 438
Db      1737 CCTCGATTGGTTTATGTTTCAAGAGAAAAACGGCCAGGCTACAACCATCACAAGAAGGCT 1796
Qy       439 GlyAlaGluAsnAlaLeuValArgValSerAlaValLeuThrAsnAlaProTyrIleLeu 458
Db      1797 GGTGCTATGAATGCATTGGTCCGAGTCTCTGCTGTACTAACTAATGCTCCTTATTTGCTG 1856
Qy       459 AsnValAspCysAspHisTyrValAsnAsnSerLysAlaValArgGluAlaMetCysIle 478
Db      1857 AACTTGGATTGTGATCACTATATCAATAATAGTAAGGCTATAAAGGAAGCAATGTGTTTT 1916
Qy       479 LeuMetAspProGlnValGlyArgAspValCysTyrValGlnPheProGlnArgPheAsp 498
Db      1917 ATGATGGATCCTTTGCTTGGAAATAAAGTTTGCTATGTGCAGTTTCCTCAAAGATTTGAT 1976
Qy       499 GlyIleAspLysSerAspArgTyrAlaAsnArgAsnValValPhePheAspValAsnMet 518
Db      1977 GGGATTGATCGCCATGATCGATATGCTAACAGAAATGTTGTCTTTTTCGATATCAACATG 2036
Qy       519 LysGlyLeuAspGlyIleGlnGlyProValTyrValGlyThrGlyCysValPheAsnArg 538
```





RESULT 3
AY103655
LOCUS AY103655 3897 bp mRNA linear HTC 16-OCT-2002
DEFINITION Zea mays PCO096398 mRNA sequence.
ACCESSION AY103655
VERSION AY103655.1 GI:21206733
KEYWORDS HTC.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 3897)
AUTHORS Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S., Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
TITLE Maize Mapping Project/DuPont Consensus Sequences for Design of Overgo Probes
JOURNAL Unpublished (2002)
REFERENCE 2 (bases 1 to 3897)
AUTHORS Coe,E.H.
TITLE Direct Submission
JOURNAL Submitted (25-APR-2002) Maize Mapping Project, University of Missouri, Columbia, MO 65211, USA
COMMENT If you are interested in getting corresponding physical clones, these are publicly available from ZmDB and may be found by BLAST searching at MSL, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR, www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the maize cDNA sequences is either Virginia Walbot, Stanford or Pat Schnable, Iowa State, then clones may be requested from ZmDB: www.zmdb.iastate.edu.
FEATURES Location/Qualifiers
source 1. .3897
/organism="Zea mays"
/mol_type="mRNA"
/db_xref="MaizeDB:635496"
/db_xref="taxon:4577"
/clone_lib="Maize Mapping Project/DuPont Cornsensus Library"
/note="this sequence is part of a project of EST assemblies resulting from the application of public contigs to seed DuPont contigs; this resource was assembled by DuPont as part of a collaboration for the overgo addressing of BACs in conjunction with the Maize Mapping Project"
BASE COUNT 982 a 839 c 1008 g 1068 t
ORIGIN

Alignment Scores:
Pred. No.: 9.25e-284 Length: 3897
Score: 3333.50 Matches: 632
Percent Similarity: 73.46% Conservative: 154

```
st Local Similarity:  59.07%        Mismatches:    175
ery Match:            63.62%        Indels:        112
                      11            Gaps:          15

-09-980-043-2 (1-978) x AY103655 (1-3897)

          3 GluSerGlyAlaProIleCysHisThrCysGlyGluGlnValGlyHisAspAlaAsnGly 22
            |||            :::|||      |||||:::   ||||||              |||
        252 GAGCAGAACGGGCAGGTGTGCCAGATTTGCGGCGACGACGTCGGCCTTGCCCCCGGCGGG 311

         23 GluLeuPheValAlaCysHisGluCysSerTyrProMetCysLysSerCysPheGluPhe 42
            :::  |||||||||||:::|||||:::::|||:::||:::  |||:::|||:::
        312 GACCCCTTCGTGGCGTGCAACGAGTGCGCCTTCCCCGTCTGCCGGGACTGCTACGAATAC 371

         43 GluIleAsnGluGlyArgLysValCysLeuArgCysGlySerProTyr------------ 58
            |||      ||||||   :::   |||   :::|||   :::   |||
        372 GAGCGCCGGGAGGGCACGCAGAACTGCCCCCAGTGCAAGACTCGATACAAGCGCCTCAAG 431

         59 --------------------------AspGluAsnLeuLeuAspAspValGluLysLys 69
                                      :::|||:::   :::||||||:::::   :::
        432 GGCTGCCAACGTGTGACCGGTGACGAGGAGGAGGACGGCGTCGATGACCTGGACAACGAG 491

         70 -- ---------GlySerGlyAsnGlnSerThrMetAlaSer- - ------  HisLeu 81
                    |||      :::||||||         |'|            ||:::
        492 TTCAACTGGGACGGCCATGACTCGCAGTCTGTGGCCGAGTCCATGCTCTACGGCCACATG 551

         82 Asn-AspSerGlnAspValGlyIleHisAlaArgHisIleSerSerValSerThrVal-- 100
            :::   :::   :::|||   :::   ||||||||      |||   ||||||
        552 AGCTACGGCCGTGGAGGTGACCCTAATGGCGCGCCACAAGCTTTCCAGCTCAACCCC-AA 610

        101 ---------------AspSerGluMetAsnAspGlu---------------------- 107
                           :::   :::|||   |||:::
        611 TGTTCCACTCCTCACCAACGGGCAAATGGTGGATGACATCCCACCGGAGCAGCACGCGCT 670

        107 ------------------------------------------------------------ 107

        671 GGTGCCTTCTTTCATGGGTGGTGGGGGAAAGAGGATACATCCCCTTCCTTATGCGGATCC 730

        108 ------------------------------------------------------------Ty 108
                                                                        ||
        731 CAGCTTACCTGTGCAACCCAGGTCTATGGACCCATCCAAGGATCTTGCTGCATATGGGTA 790

        108 rGlyAsnProIleTrpLysAsnArgValLysSerCysLysAspLysGluAsnLysLysLy 128
            |||:::     |||||   |||:::::::::   |||   ::::::   :::
        791 TGGTAGTGTTGCTTGGAAGGAACGGATGGAGAATTGGAAGCAGAGACAAGAGAGGATGCA 850

        128 sLysArgSerProLysAla---------GluThrGluProAlaGlnValProThrGluGl 145
             :::                      :::   :::   |||   :::|||
        851 CCAGACGGGGAATGATGGTGGTGGTGATGATGGTGACGATGCTGATCTACCACTAATGGA 910

        145 nGlnMetGluGluLysProSerAlaGluAlaSerGluProLeuSerIleValTyrProIl 165
                                    ||||||   :::   ||||||          |||::
        911 T-----------------------GAAGCAAGACAACAACTGTCCAGGAAAATTCCACT 946

        165 eProArgAsnLysLeuThrProTyrArgAlaValIleIleMetArgLeuValIleLeuGl 185
            :|||   :::::::::   |||||||||   :::||||||:::|||||||||:::||||||
)       947 TCCATCAAGCCAGATTAATCCATATAGGATGATTATCATTATTCGGCTTGTGGTTTTGGG 1006

/       185 yLeuPhePheHisPheArgIleThrAsnProValAspSerAlaPheGlyLeuTrpLeuTh 205
            |   |||||||||:::|||:::   :::|||||||:::   ||||||   ||||||
>      1007 GTTCTTCTTCCACTACCGAGTGATGCATCCGGTGAATGATGCATTTGCTTTGTGGCGCAT 1066

/       205 rSerValIleCysGluIleTrpPheAlaPheSerTrpValLeuAspGlnPheProLysTr 225
            ||||||||||||||||||||||||||||||   ||||||:::||||||||||||||||||||
>      1067 ATCTGTTATCTGTGAAATCTGGTTTGCCATGTCTTGGATTCTTGATCAATTCCCAAAGTG 1126

/       225 pAsnProValAsnArgGluThrTyrIleGluArgLeuSerAlaArgTyrGluArgGluGl 245
            |   |||:::   |||||||||||:::::::||||||||   |||:::::::::||||||
>      1127 GTTCCCTATTGAGAGAGAGACTTACCTAGACCGGCTGTCACTGAGGTTCGACAAGGAAGG 1186

/       245 yGluProSerGlnLeuAlaGlyValAspPhePheValSerThrValAspProLeuLysGl 265
            |:::|||||||||||||||   :::|||||||||||||||||||||||||||||||||||
>      1187 CCAGCCATCTCAACTTGCTCCAATTGATTTCTTTGTCAGTACGGTTGATCCCTTAAAGGA 1246
```

```
Qy       265 uProProLeuIleTh::AlaAsnThrValLeuSerIleLeuAlaValAspTyrProValAs 285
             ||||||||||:::||  ||  ||||:|||| ||||||:::|||||||||||||||:||
Db      1247 ACCTCCTTTGGTCA AACAAATACTGTTCTATCTATCCTTTCGGTGGATTATCCTGTTGA 1306

Qy       285 pLysValSerCysTy::ValSerAspAspGlyAlaAlaMetLeuSerPheGluSerLeuVa 305
             |||||:||||||  ||| :|||: ||| :||||:|||||'|:::|||||:::|||
Db      1307 TAAGGTTTCTTGCTATGTTTCTGATGATGGTGCTGCAATGCTAACGTTTGAAGCATTATC 1366

Qy       305 lGluThrAlaGluPheAlaArgLysTrpValProPheCysLysLysPheSerIleGluPr 325
             ||:|||:::||||  ||||:::||| .|||||  |||||:|||||||||:::::::::|||| ||
Db      1367 TGAAACATCTGAATT:TGCAAAGAAATGGGTTCCTTTCTGCAAACGGTACAATATTGAACC 1426

Qy       325 oArgAlaProGluPh:eTyrPheSerGlnLysIleAspTyrLeuLysAspLysValGlnPr 345
             ||||||'|||: |:::|||||    |:|||||.|||||||||||||||||||'|'||
Db      1427 TCGCGCTCCAGAGT'TTACTTCCAACAGAAGATAGACTACTTGAAAGACAAGGTGGCAGC 1486

Qy       345 oSerPheValLysGlnArgArgAlaMetLysArgAspTyrGluGluTyrLysValArgVa 365
             :::|||: |:::::|. |||  |||.|'|||:::|||||||:::|||||'||::
Db      1487 AAACTTTGTTAG'''''AAGAGAAATGAAGAGAGAGTATGAGGAATTCAAGGTGAGAAT 1546

Qy       365 lAsnAlaLeuVal:::::::::::::::::::::::::::::::::::::::::::::Gl 385
             :|||:::|:: |:  :|  |  |  :|||:|||||||||||||||||||||||||||||
Db      1547 CAATGCTTAGT'':  AAAG'''A AAAGTTC TGAAGAAGGATGGACAATGCAAGATGG 1606

Qy       385 yThrProTrpProGlyAsnAsnThrArgAspHisProGlyHisAspSerGlyLeuProTr 405
             |||||.|||:|||.||||:   |.||||. |||||   |.|||||.|||||||||||||
Db      1607 AACCCCCTGGCCTGGAAACAATGTTCGTGATCATCCTGGAAT GATTCAGGTCTTCCTTG 1666

Qy       405 pGluIleLeuGlyAlaArgAspIleGluGlyAsnGluLeuProArgLeuValTyrValSe 425
             |   ||      |||  ||||:|||.|||||||||||||||:|||||||||||||||||
Db      1666 GCCAA--AGCGGAGCCCTTGACTGTGAGGGAAATGAACTGCCACGATTGGTTTATGTTTC 1723

Qy       425 rArgGluLysArgPr:GlyTyrGlnHisHisLysLysAlaGlyAlaGluAsnAlaLeuVa 445
             ||||.|||.||||'''|||'|   :'|||||'||||||||||||||   |.|||||||||
Db      1724 TAGAGAGAAACGAC'AGGCTATAACCATCATAAGAAAGCTGGTGCTATGAATGCATTGGT 1783

Qy       445 lArgValSerAlaValLeuThrAsnAlaProTyrIleLeuAsnValAspCysAspHisTy 465
             |||||||||||||||'||||||'||||'|||||':::||||||:::|||||||||||||||
Db      1784 CCGAGTCTCTGCTGTACTAACAAATGCTCCATATTTGTTAAACTTGGATTGTGATCACTA 1843

Qy       465 rValAsnAsnSerLysAlaValArgGluAlaMetCysIleLeuMetAspProGlnValGl 485
             |:::|||||||||||||:::::||||||||||.|   :::||||||||   :::||
Db      1844 CATCAACAACAGCAAGGCTATAAAGGAAGCAATGTGTTTTATGATGGACCCTTTACTAGG 1903

Qy       485 yArgAspValCysTyrValGlnPheProGlnArgPheAspGlyIleAspLysSerAspAr 505
             |:::  ||.||||.||||||||.|||||.||||||||||||||:::   ||||
Db      1904 AAAGAAGGTTTGCTATGTACAGTTCCCTCAAAGATTTGATGGGATTGATCGCCATGACCG 1963

Qy       505 gTyrAlaAsnArgAsnValVal:PhePheAspValAsnMetLysGlyLeuAspGlyIleGl 525
             ||||||||||'|||'||||||||||||:|||:::|||||||||||||||||:|||||||
Db      1964 ATATGCTAACCGGAATGTTGTCTTTTTTGATATCAACATGAAAGGTTTGGATGGTATTCA 2023

Qy       525 nGlyProValTyrValGlyThrGlyCysValPheAsnArgGlnAlaLeuTyrGlyTyrGl 545
             ||,|||:::|'||||||||||||||||||||   |||||||||||||||||||||||||
Db      2024 GGGTCCAATTTATGTTGGTACTGGATGTGTATTTAGAAGGCAGGCATTATATGGTTATGA 2083

Qy       545 yProProSerMet:ProSerLeuArgLysArgLysAspSerSerSer---------- --- 560
                 |||              |||:::   |||   :::
Db      2084 TGCCCCCAAA----  -----ACAAAGAAGCCACCATCAAGGACTTGCAACTGCTGGCC 2131

Qy       561 ----  CysPheSer:CysCysCys  ----ProSerLysLysLysProAlaGlnAspPr 576
                   |':|||   |||||||||              |||:::|||          :|
Db      2132 CAAGTGGTGCTTTTGCTGTTGCTGCTTTGGCAATAGGAAGCAAAAGAAGACTACCAAACC 2191

Qy       576 oAlaGluValTyrArgAspAla---------LysArgGluAspLeuAsnAlaAlaIlePh 593
             |              :::                |||:::|||:::   :::   ::
Db      2192 CAAAACAGAGAAGAAAAAGTTATTATTTTTCAAGAAAGAAGAGAACCAATCCCCTGCATA 2251

Qy       593 eAsnLeuThrGluIleAspAsnTyr---------AspGluHisGluArgSerMetLeuIl 610
             :   |||   |||||||||                |||:::|||::::::   :::::
Db      2252 TGCTCTTGGTGAAATTGACGAAGCTGCTCCAGGAGCTGAGAATGAAAAGGCCGGTATTGT 2311
```

```
Qy       610 eSerGlnLeuSerPheGluLysThrPheGlyLeuSerSerValPheIleGluSerThrLe 630
Db      2312 AAATCAACAAAAATTAGAAAAGAAATTTGGCCAATCTTCTGTTTTTGTTACATCCACACT 2371

Qy       630 uMetGluAsnGlyGlyValProGluSerAlaAsnSerProProPheIleLysGluAlaIl 650
Db      2372 TCTCGAGAATGGTGGAACCTTGAAGAGTGCAAGTCCTGCTTCTCTTTTGAAAGAAGCTAT 2431

Qy       650 eGlnValIleGlyCysGlyTyrGluGluLysThrGluTrpGlyLysGlnIleGlyTrpIl 670
Db      2432 ACATGTCATTAGTTGTGGTTATGAAGACAAGACAGACTGGGGAAAAGAGATTGGCTGGAT 2491

Qy       670 eTyrGlySerValThrGluAspIleLeuSerGlyPheLysMetHisCysArgGlyTrpAr 690
Db      2492 CTATGGATCAGTTACAGAAGATATTCTAACTGGTTTCAAGATGCATTGTCATGGTTGGCG 2551

Qy       690 gSerIleTyrCysMetProValArgProAlaPheLysGlySerAlaProIleAsnLeuSe 710
Db      2552 GTCAATTTACTGCATACCTAAACGGGTTGCATTCAAAGGTTCTGCACCTCTGAATCTTTC 2611

Qy       710 rAspArgLeuHisGlnValLeuArgTrpAlaLeuGlySerValGluIlePhePheSerAr 730
Db      2612 AGATCGTCTTCACCAGGTGCTTCGGTGGGCTCTTGGGTCTATTGAGATCTTCTTCAGCAA 2671

Qy       730 gHisCysProLeuTrpTyrGlyPheGlyGlyGlyArgLeuLysTrpLeuGlnArgLeuAl 750
Db      2672 TCATTGCCCTCTTTGGTATGGGTATGGTGGCGGT---CTGAAATTTTTGGAAAGATTTTC 2728

Qy       750 aTyrIleAsnThrIleValTyrProPheThrSerLeuProLeuIleAlaTyrCysThrIl 770
Db      2729 CTACATCAACTCCATCGTGTATCCTTGGACATCTATTCCCCTCTTGGCTTACTGTACATT 2788

Qy       770 eProAlaValCysLeuLeuThrGlyLysPheIleIleProThrLeuSerAsnLeuAlaSe 790
Db      2789 GCCTGCCATCTGTTTATTGACAGGGAAATTTATCACTCCAGAGCTGAATAATGTTGCCAG 2848

Qy       790 rMetLeuPheLeuGlyLeuPheIleSerIleIleValThrAlaValLeuGluLeuArgTr 810
Db      2849 CCTGTGGTTCATGTCACTTTTTATCTGCATTTTTGCTACGAGCATCCTAGAAATGAGATG 2908

Qy       810 pSerGlyValSerIleGluAspLeuTrpArgAsnGluGlnPheTrpValIleGlyGlyVa 830
Db      2909 GAGTGGTGTTGGAATTGATGACTGGTGGAGGAATGAGCAGTTCTGGGTCATTGGAGGTGT 2968

Qy       830 lSerAlaHisLeuPheAlaValPheGlnGlyPheLeuLysMetLeuAlaGlyIleAspTh 850
Db      2969 GTCCTCACACCTCTTTGCTGTGTTCCAGGGACTTCTCAAGCTATAGCTGGTGATGATAC 3028

Qy       850 rAsnPheThrValThrAlaLysAlaAlaGluAspAlaLeuGlyGluLeuTyrMetVa 870
Db      3029 AAGCTTCACCGTGACATCAAAGGGTGGAGATGATGAGGAGTTCTCAGAGCTATATACATT 3088

Qy       870 lLysTrpThrThrLeuLeuIleProProThrThrLeuLeuIleIleAsnMetSerGly-- 889
Db      3089 CAAATGGACTACCTTATTGATACCTCCTACCACCTTGCTTCTATTGAACTTCATTGGTGT 3148

Qy       890 -CysAlaGlyPheSerAspAlaLeuAsnLysGlyTyrGluAlaTrpGlyProLeuPheGl 909
Db      3149 GGTCGCTGGCGTTTCAAATGCGATCAATAACGGATATGAGTTATGGGGCCCCCTCTTTGG 3208

Qy       909 yLysValPhePheAlaPheTrpValIleLeuHisLeuTyrProPheLeuLysGlyLeuMe 929
Db      3209 GAAGCTATTCTTTGCATTTTGGGTGATTGTCCATCTTTATCCTTTCTCAAAGGTTTGGT 3268

Qy       929 tGlyArgGlnAsnLeuThrProThrIleValValLeuTrpSerValLeuLeuAlaSerVa 949
Db      3269 TGGAAGGCAAAACAGGACACCAACGATTGTCATCGTCTGGTCCATTCTGCTGGCTTCAAT 3328

Qy       949 lPheSerLeuValTrpValLysIleAsnProPheValAsnLysValAspAsnThrLeuVa 969
Db      3329 CTTCTCGCTCCTTTGGGTTCGGATTGATCCTTTCCTTGCGAAGGATGATGGTCCGCTTCT 3388

Qy       969 lAlaGluThrCysIleSerIleAspCys 978
```

```
Db      3389 TGAGGAG------TGTGGTTTGGATTGC 3410

RESULT 4
AY103701
LOCUS       AY103701                3788 bp    mRNA    linear   HTC 16-OCT-2002
DEFINITION  Zea mays PCO120363 mRNA sequence.
ACCESSION   AY103701
VERSION     AY103701.1  GI:21206779
KEYWORDS    HTC.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 3788)
  AUTHORS   Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S.,
            Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
  TITLE     Maize Mapping Project/DuPont Consensus Sequences for Design of
            Overgo Probes
  JOURNAL   Unpublished (2002)
REFERENCE   2  (bases 1 to 3788)
  AUTHORS   Coe,E.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-APR-2002) Maize Mapping Project, University of
            Missouri, Columbia, MO 65211, USA
COMMENT     If you are interested in getting corresponding physical clones,
            these are publicly available from ZmDB and may be found by BLAST
            searching at MSL, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR,
            www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the
            maize cDNA sequences is either Virginia Walbot, Stanford or Pat
            Schnable, Iowa State, then clones may be requested from ZmDB:
            www.zmdb.iastate.edu.
FEATURES             Location/Qualifiers
     source          1. .3788
                     /organism="Zea mays"
                     /mol_type="mRNA"
                     /db_xref="MaizeDB:637005"
                     /db_xref="taxon:4577"
                     /clone_lib="Maize Mapping Project/DuPont Cornsensus
                     Library"
                     /note="this sequence is part of a project of EST
                     assemblies resulting from the application of public
                     contigs to seed DuPont contigs; this resource was
                     assembled by DuPont as part of a collaboration for the
                     overgo addressing of BACs in conjunction with the Maize
                     Mapping Project"
BASE COUNT      851 a    954 c   1114 g    866 t      3 others
ORIGIN

Alignment Scores:
Pred. No.:                  3.29e-281       Length:          3788
Score:                      3304.50         Matches:         631
Percent Similarity:         72.53%          Conservative:    148
Best Local Similarity:      58.75%          Mismatches:      176
Query Match:                63.06%          Indels:          121
DB:                         11              Gaps:            15

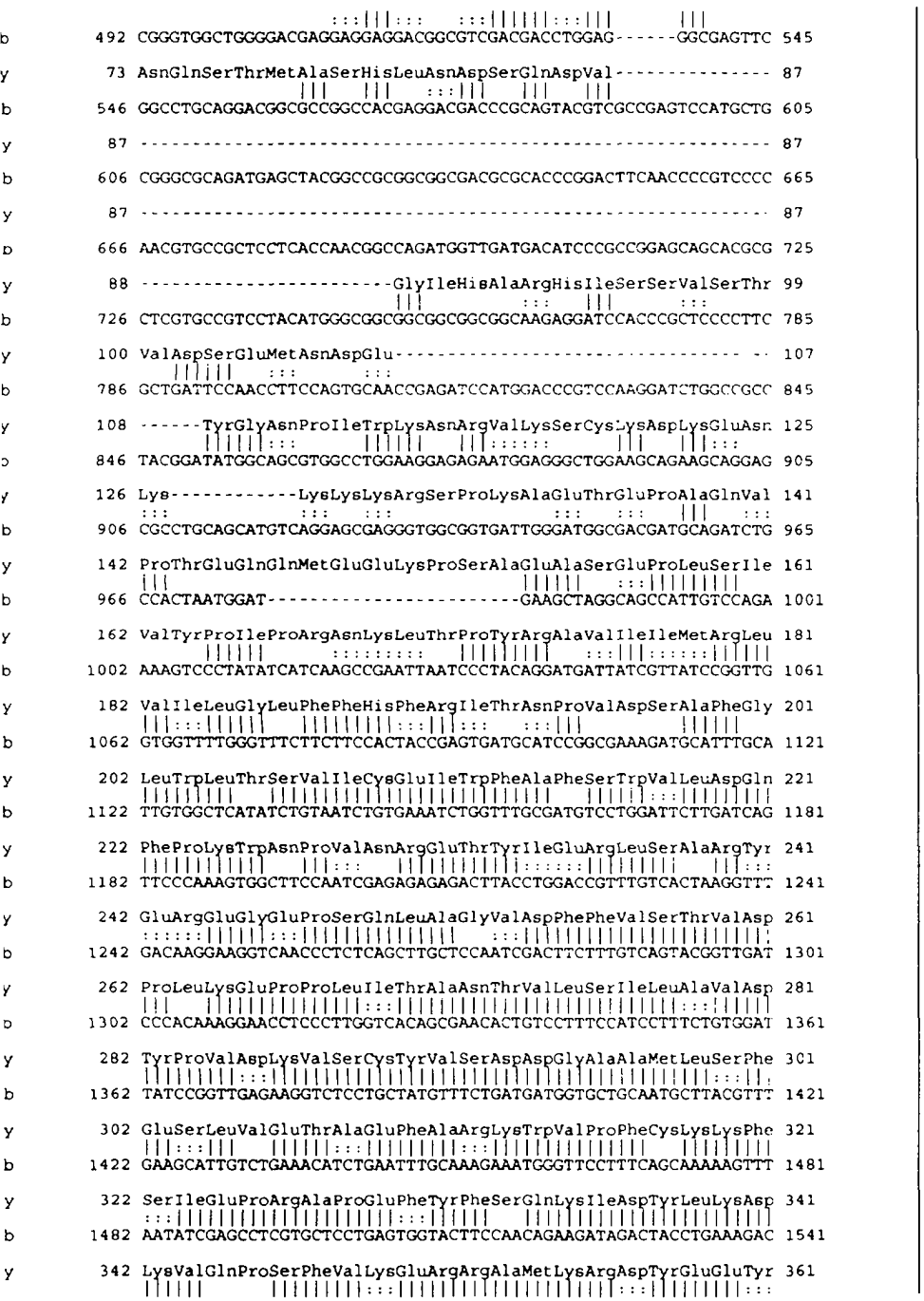
US-09-980-043-2 (1-978) x AY103701 (1-3788)

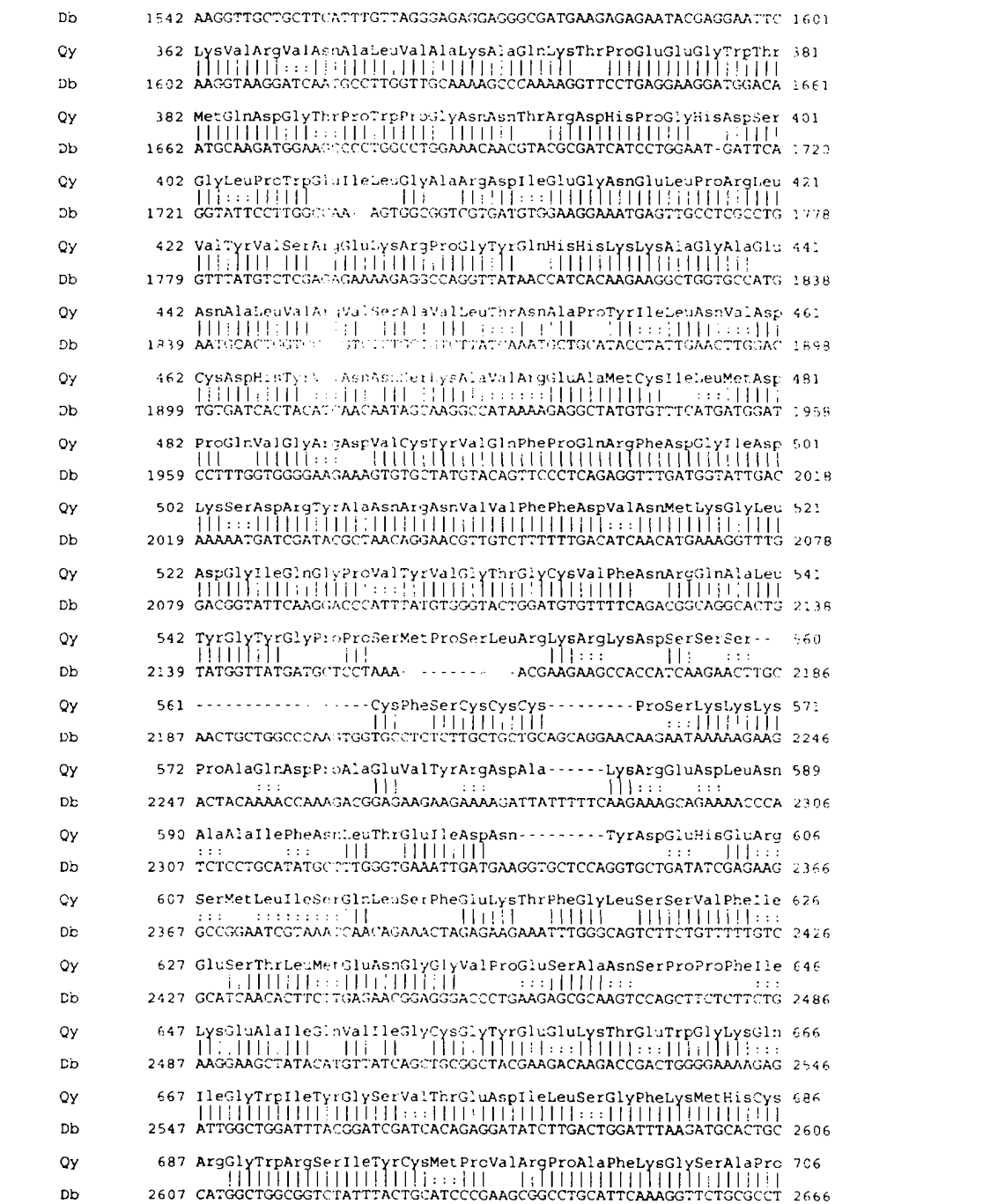
Qy         6 AlaProIleCysHisThrCysGlyGluGlnValGlyHisAspAlaAsnGlyGluLeuPhe 25
Db       315 GCGCCG---TGCCAGATATGCGGCGACGAGGTCGGGGTGGGCTTCGACGGGGAGCCCTTC 371

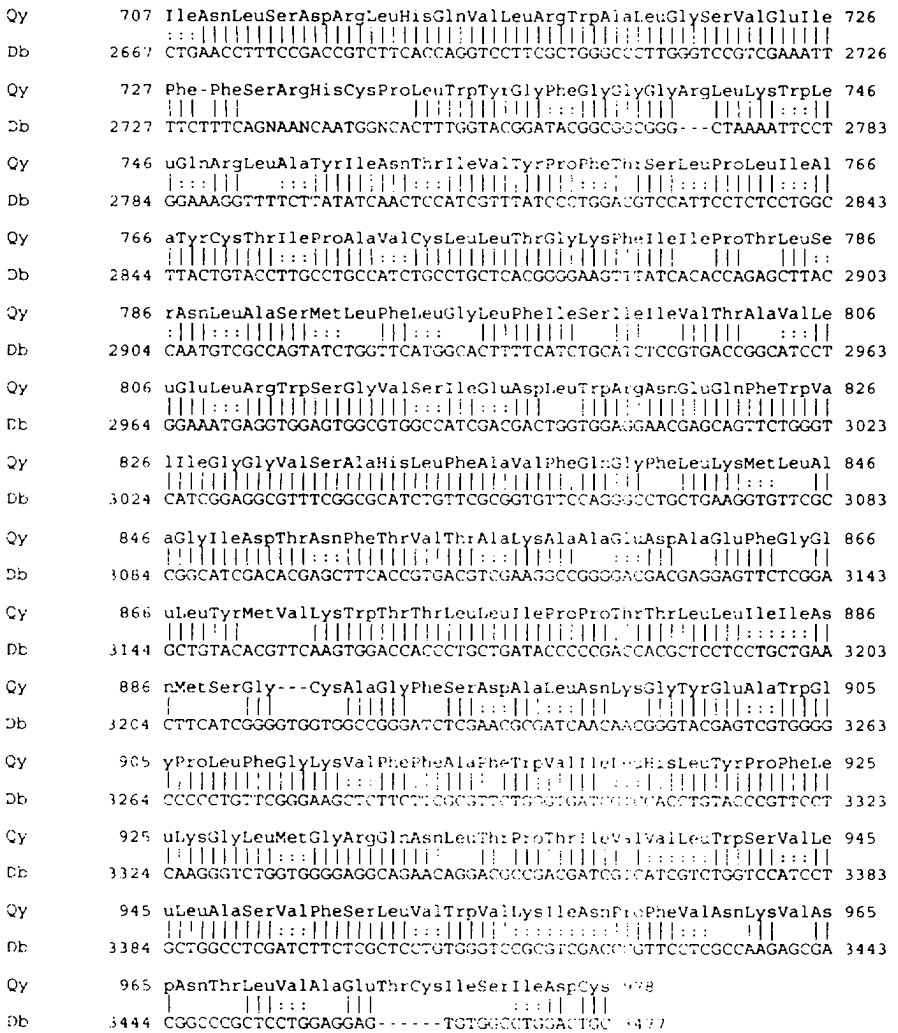
Qy        26 ValAlaCysHisGluCysSerTyrProMetCysLysSerCysPheGluPheGluIleAsn 45
Db       372 GTGGCGTGCAACGAGTGCGCCTTCCCCGTCTGCCGCGCCTGCTACGAGTACGAGCGCCGC 431

Qy        46 GluGlyArgLysValCysLeuArgCysGlySerProTyr---------------------- 58
Db       432 GAGGGCTCGCAGGCGTGCCCGCAGTGCAGGACCCGCTACAAGCGCCTCAAGGGTTGCCCG 491

Qy        59 ------------------AspGluAsnLeuLeuAspAspValGluLysLysGlySerGly 72
```


b 492 CGGGTGGCTGGGGACGAGGAGGAGGACGGCGTCGACGACCTGGAG------GGCGAGTTC 545
y 73 AsnGlnSerThrMetAlaSerHisLeuAsnAspSerGlnAspVal---------- 87
b 546 GGCCTGCAGGACGGCGCCGGCCACGAGGACGACCCGCAGTACGTCGCCGAGTCCATGCTG 605
y 87 ---------- 87
b 606 CGGGCGCAGATGAGCTACGGCCGCGGCGGCGACGCGCACCCGGACTTCAACCCCGTCCCC 665
y 87 ---------- 87
b 666 AACGTGCCGCTCCTCACCAACGGCCAGATGGTTGATGACATCCCGCCGGAGCAGCACGCG 725
y 88 ----------GlyIleHisAlaArgHisIleSerSerValSerThr 99
b 726 CTCGTGCCGTCCTACATGGGCGGCGGCGGCGGCGGCAAGAGGATCCACCCGCTCCCCTTC 785
y 100 ValAspSerGluMetAsnAspGlu---------- 107
b 786 GCTGATTCCAACCTTCCAGTGCAACCGAGATCCATGGACCCGTCCAAGGATCTGGCCGCC 845
y 108 ------TyrGlyAsnProIleTrpLysAsnArgValLysSerCysLysAspLysGluAsn 125
b 846 TACGGATATGGCAGCGTGGCCTGGAAGGAGAGAATGGAGGGCTGGAAGCAGAAGCAGGAG 905
y 126 Lys----------LysLysLysArgSerProLysAlaGluThrGluProAlaGlnVal 141
b 906 CGCCTGCAGCATGTCAGGAGCGAGGGTGGCGGTGATTGGGATGGCGACGATGCAGATCTG 965
y 142 ProThrGluGlnGlnMetGluGluLysProSerAlaGluAlaSerGluProLeuSerIle 161
b 966 CCACTAATGGAT----------GAAGCTAGGCAGCCATTGTCCAGA 1001
y 162 ValTyrProIleProArgAsnLysLeuThrProTyrArgAlaValIleIleMetArgLeu 181
b 1002 AAAGTCCCTATATCATCAAGCCGAATTAATCCCTACAGGATGATTATCGTTATCCGGTTG 1061
y 182 ValIleLeuGlyLeuPhePheHisPheArgIleThrAsnProValAspSerAlaPheGly 201
b 1062 GTGGTTTTGGGTTTCTTCTTCCACTACCGAGTGATGCATCCGGCGAAAGATGCATTTGCA 1121
y 202 LeuTrpLeuThrSerValIleCysGluIleTrpPheAlaPheSerTrpValLeuAspGln 221
b 1122 TTGTGGCTCATATCTGTAATCTGTGAAATCTGGTTTGCGATGTCCTGGATTCTTGATCAG 1181
y 222 PheProLysTrpAsnProValAsnArgGluThrTyrIleGluArgLeuSerAlaArgTyr 241
b 1182 TTCCCAAAGTGGCTTCCAATCGAGAGAGAGACTTACCTGGACCGTTTGTCACTAAGGTTT 1241
y 242 GluArgGluGlyGluProSerGlnLeuAlaGlyValAspPhePheValSerThrValAsp 261
b 1242 GACAAGGAAGGTCAACCCTCTCAGCTTGCTCCAATCGACTTCTTTGTCAGTACGGTTGAT 1301
y 262 ProLeuLysGluProProLeuIleThrAlaAsnThrValLeuSerIleLeuAlaValAsp 281
b 1302 CCCACAAAGGAACCTCCCTTGGTCACAGCGAACACTGTCCTTTCCATCCTTTCTGTGGAT 1361
y 282 TyrProValAspLysValSerCysTyrValSerAspAspGlyAlaAlaMetLeuSerPhe 301
b 1362 TATCCGGTTGAGAAGGTCTCCTGCTATGTTTCTGATGATGGTGCTGCAATGCTTACGTTT 1421
y 302 GluSerLeuValGluThrAlaGluPheAlaArgLysTrpValProPheCysLysLysPhe 321
b 1422 GAAGCATTGTCTGAAACATCTGAATTTGCAAAGAAATGGGTTCCTTTCAGCAAAAAGTTT 1481
y 322 SerIleGluProArgAlaProGluPheTyrPheSerGlnLysIleAspTyrLeuLysAsp 341
b 1482 AATATCGAGCCTCGTGCTCCTGAGTGGTACTTCCAACAGAAGATAGACTACCTGAAAGAC 1541
y 342 LysValGlnProSerPheValLysGluArgArgAlaMetLysArgAspTyrGluGluTyr 361


Db 1542 AAGGTTGCTGCTTCATTTGTTAGGGAGAGGAGGGCGATGAAGAGAGAATACGAGGAATTC 1601
Qy 362 LysValArgValAsnAlaLeuValAlaLysAlaGlnLysThrProGluGluGlyTrpThr 381
Db 1602 AAGGTAAGGATCAATGCCTTGGTTGCAAAAGCCCAAAAGGTTCCTGAGGAAGGATGGACA 1661
Qy 382 MetGlnAspGlyThrProTrpProGlyAsnAsnThrArgAspHisProGlyHisAspSer 401
Db 1662 ATGCAAGATGGAAGCCCTGGCCTGGAAACAACGTACGCGATCATCCTGGAAT-GATTCA 1720
Qy 402 GlyLeuProTrpGluIleLeuGlyAlaArgAspIleGluGlyAsnGluLeuProArgLeu 421
Db 1721 GGTATTCCTTGGCCAA- AGTGGCGGTCGTGATGTGGAAGGAAATGAGTTGCCTCGCCTG 1778
Qy 422 ValTyrValSerArgGluLysArgProGlyTyrGlnHisHisLysLysAlaGlyAlaGlu 441
Db 1779 GTTTATGTCTCGAGAGAAAAGAGGCCAGGTTATAACCATCACAAGAAGGCTGGTGCCATG 1838
Qy 442 AsnAlaLeuValArgValSerAlaValLeuThrAsnAlaProTyrIleLeuAsnValAsp 461
Db 1839 AATGCACTGGTCC- GTCCTCTCCTTTTATCAAATGCTGCATACCTATTGAACTTGGAC 1898
Qy 462 CysAspHisTyr- -AspAsnGlnLysAlaValArgGluAlaMetCysIleLeuMetAsp 481
Db 1899 TGTGATCACTACAT-AACAATAGCAAGGCCATAAAAGAGGCTATGTGTTTCATGATGGAT 1958
Qy 482 ProGlnValGlyArgAspValCysTyrValGlnPheProGlnArgPheAspGlyIleAsp 501
Db 1959 CCTTTGGTGGGGAAGAAAGTGTGCTATGTACAGTTCCCTCAGAGGTTTGATGGTATTGAC 2018
Qy 502 LysSerAspArgTyrAlaAsnArgAsnValValPhePheAspValAsnMetLysGlyLeu 521
Db 2019 AAAAATGATCGATACGCTAACAGGAACGTTGTCTTTTTTGACATCAACATGAAAGGTTTG 2078
Qy 522 AspGlyIleGlnGlyProValTyrValGlyThrGlyCysValPheAsnArgGlnAlaLeu 541
Db 2079 GACGGTATTCAAGGACCCATTTATGTGGGTACTGGATGTGTTTTCAGACGGCAGGCACTG 2138
Qy 542 TyrGlyTyrGlyProProSerMetProSerLeuArgLysArgLysAspSerSerSer-- 560
Db 2139 TATGGTTATGATGCTCCTAAA- ------- -ACGAAGAAGCCACCATCAAGAACTTGC 2186
Qy 561 ---------- -----CysPheSerCysCysCys---------ProSerLysLysLys 571
Db 2187 AACTGCTGGCCCAAAGTGGTGCCTCTCTTGCTGCTGCAGCAGGAACAAGAATAAAAAGAAG 2246
Qy 572 ProAlaGlnAspProAlaGluValTyrArgAspAla------LysArgGluAspLeuAsn 589
Db 2247 ACTACAAAACCAAAGACGGAGAAGAAGAAAAGATTATTTTTCAAGAAAGCAGAAAACCCA 2306
Qy 590 AlaAlaIlePheAsnLeuThrGluIleAspAsn---------TyrAspGluHisGluArg 606
Db 2307 TCTCCTGCATATGCTTTGGGTGAAATTGATGAAGGTGCTCCAGGTGCTGATATCGAGAAG 2366
Qy 607 SerMetLeuIleSerGlnLeuSerPheGluLysThrPheGlyLeuSerSerValPheIle 626
Db 2367 GCCGGAATCGTAAATCAACAGAAACTAGAGAAGAAATTTGGGCAGTCTTCTGTTTTTGTC 2426
Qy 627 GluSerThrLeuMetGluAsnGlyGlyValProGluSerAlaAsnSerProProPheIle 646
Db 2427 GCATCAACACTTCTTGAGAACGGAGGGACCCTGAAGAGCGCAAGTCCAGCTTCTCTTCTG 2486
Qy 647 LysGluAlaIleGlnValIleGlyCysGlyTyrGluGluLysThrGluTrpGlyLysGln 666
Db 2487 AAGGAAGCTATACATGTTATCAGCTGCGGCTACGAAGACAAGACCGACTGGGGAAAAGAG 2546
Qy 667 IleGlyTrpIleTyrGlySerValThrGluAspIleLeuSerGlyPheLysMetHisCys 686
Db 2547 ATTGGCTGGATTTACGGATCGATCACAGAGGATATCTTGACTGGATTTAAGATGCACTGC 2606
Qy 687 ArgGlyTrpArgSerIleTyrCysMetProValArgProAlaPheLysGlySerAlaPro 706
Db 2607 CATGGCTGGCGGTCTATTTACTGCATCCCGAAGCGGCCTGCATTCAAAGGTTCTGCGCCT 2666

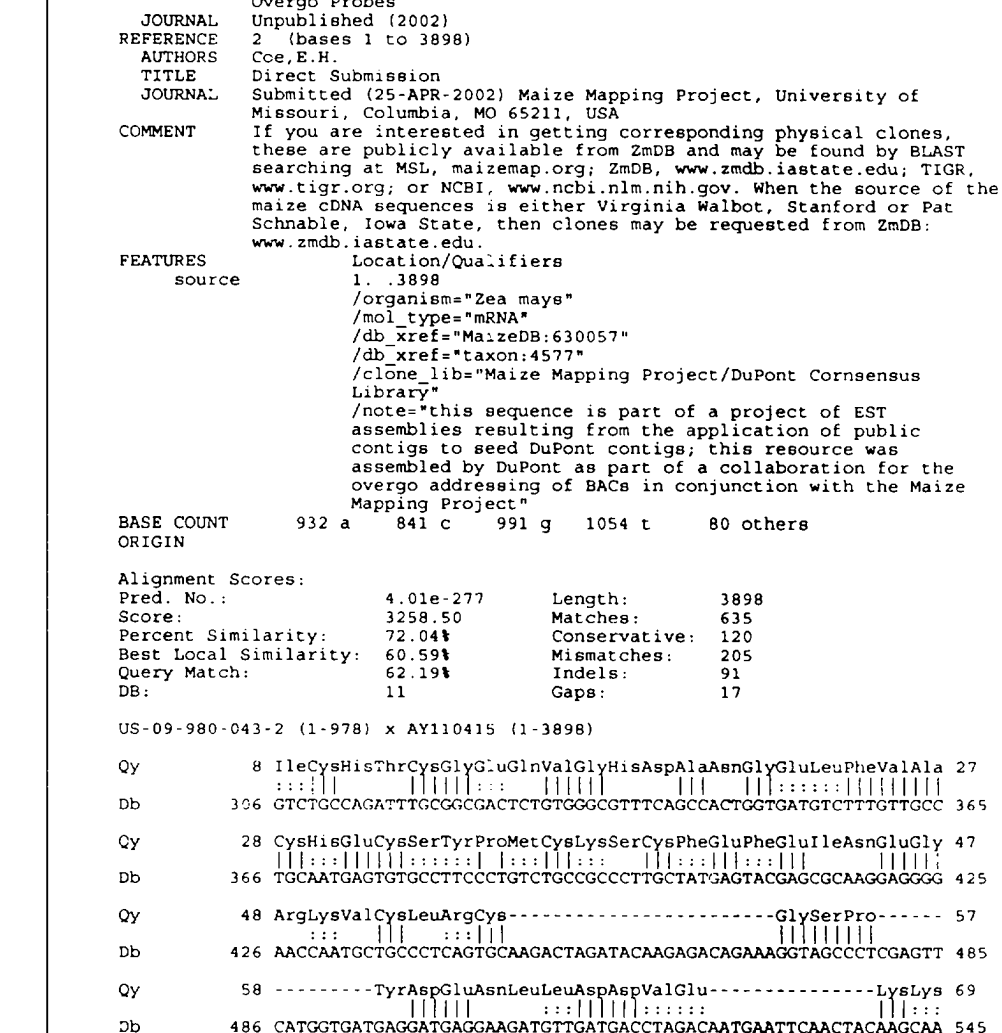


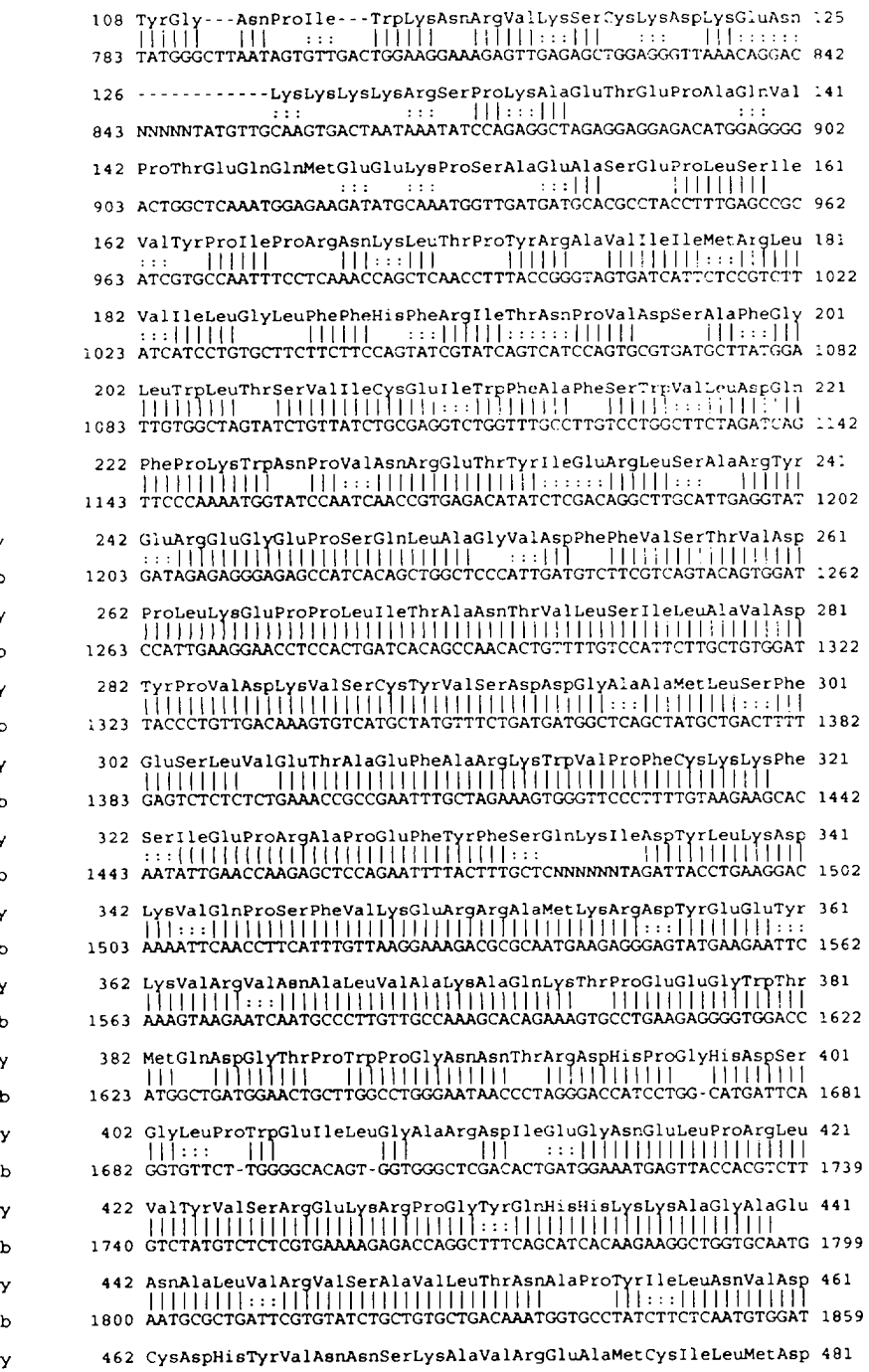
RESULT 5
AY110415
LOCUS AY110415 3898 bp mRNA linear HTC 17-OCT-2002
DEFINITION Zea mays CL1166_1 mRNA sequence.
ACCESSION AY110415
VERSION AY110415.1 GI:21214824
KEYWORDS HTC.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 3898)
AUTHORS Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S., Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
TITLE Maize Mapping Project/DuPont Consensus Sequences for Design of Overgo Probes
JOURNAL Unpublished (2002)
REFERENCE 2 (bases 1 to 3898)
AUTHORS Coe,E.H.
TITLE Direct Submission
JOURNAL Submitted (25-APR-2002) Maize Mapping Project, University of Missouri, Columbia, MO 65211, USA
COMMENT If you are interested in getting corresponding physical clones, these are publicly available from ZmDB and may be found by BLAST searching at MSL, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR, www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the maize cDNA sequences is either Virginia Walbot, Stanford or Pat Schnable, Iowa State, then clones may be requested from ZmDB: www.zmdb.iastate.edu.
FEATURES Location/Qualifiers
source 1. .3898
/organism="Zea mays"
/mol_type="mRNA"
/db_xref="MaizeDB:630057"
/db_xref="taxon:4577"
/clone_lib="Maize Mapping Project/DuPont Cornsensus Library"
/note="this sequence is part of a project of EST assemblies resulting from the application of public contigs to seed DuPont contigs; this resource was assembled by DuPont as part of a collaboration for the overgo addressing of BACs in conjunction with the Maize Mapping Project"
BASE COUNT 932 a 841 c 991 g 1054 t 80 others
ORIGIN

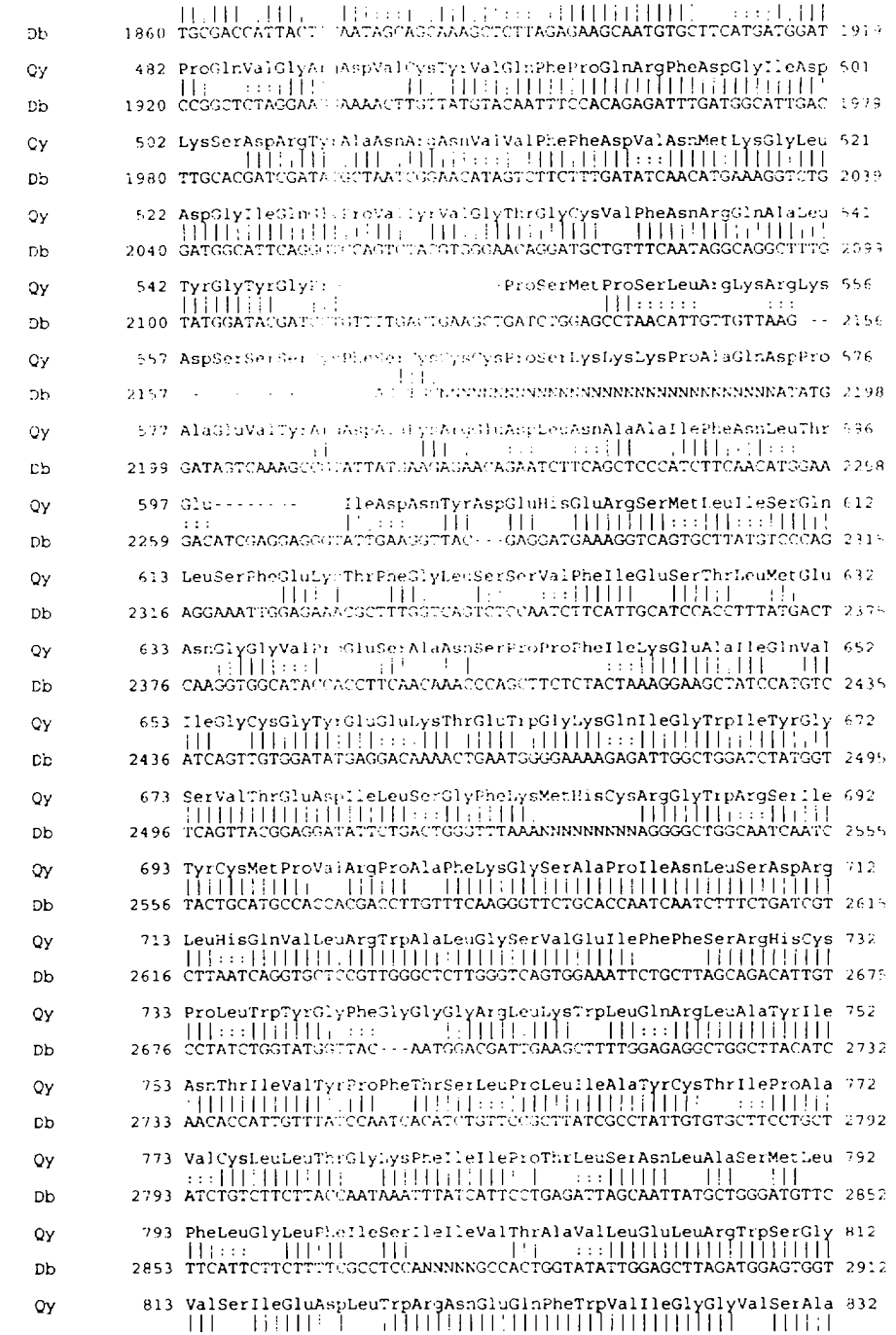
Alignment Scores:

Pred. No.:	4.01e-277	Length:	3898
Score:	3258.50	Matches:	635
Percent Similarity:	72.04%	Conservative:	120
Best Local Similarity:	60.59%	Mismatches:	205
Query Match:	62.19%	Indels:	91
DB:	11	Gaps:	17

US-09-980-043-2 (1-978) x AY110415 (1-3898)







```
Db       2913 GTTGGCATTGAAGATTGGTGGAGAAATGAGCAGTTTTGGGTTATTGGTGGCACCTCTGCC 2972

Qy        833 HisLeuPheAlaValPheGlnGlyPheLeuLysMetLeuAlaGlyIleAspThrAsnPhe 852

Db       2973 CATCTCTTCGCAGTGTTCCAGGGTCTGCTGAAAGTGTTGGCTGGGATTGATACCAACTTC 3032

Qy        853 ThrValThrAlaLysAlaAla-GluAspAlaGluPheGlyGluLeuTyrMetValLys 871

Db       3033 ACAGTTACCTCAAAGGCATCTGATGAGGATGGCGACTTTGCTGAGCTATATGTGTTCAAG 3092

Qy        872 TrpThrThrLeuLeuIleProProThrThrLeuLeuIleIleAsnMetSerGly---Cys 890

Db       3093 TGGACCAGTTTGCTCATTCCTCCGACCACTGTTCTTGTCATTAACCTGGTCGGAATGGTG 3152

Qy        891 AlaGlyPheSerAspAlaLeuAsnLysGlyTyrGluAlaTrpGlyProLeuPheGlyLys 910

Db       3153 GCAGGAATTTCGTATGCCATTAACAGTGGCTACCAATCCTGAGGTCCGCTCTTTGGAAAG 3212

Qy        911 ValPhePheAlaPheTrpValIleLeuHisLeuTyrProPheLeuLysGlyLeuMetGly 930

Db       3213 CTGTTCTTCTCGATCTGGGTGNTCCTCCATCTGTACACGTTTCTCAAGGGTCTCATGGGA 3272

Qy        931 ArgGlnAsnLeuThrProThrIleValValLeuTrpSerValLeuLeuAlaSerValPhe 950

Db       3273 AGGCAGAACCGCACACCAACAATCGTCATTGTCTGGTCCATTCTTCTTGCATCTATCTTC 3332

Qy        951 SerLeuValTrpValLysIleAsnProPheValAsnLysValAspAsnThrLeuValAla 970

Db       3333 TCCTTGCTGTGGGTGAAGATCGATCCTTTCATCTCCCCGACACAGAAAGCTGCTGCCTTG 3392

Qy        971 GluThrCysIleSerIleAspCys 978

Db       3393 GGGCAATGT---GGCGTCAACTGC 3413
```

RESULT 6
AY104236

```
LOCUS       AY104236                2872 bp    mRNA    linear   HTC 16-OCT-2002
DEFINITION  Zea mays PCO121439 mRNA sequence.
ACCESSION   AY104236
VERSION     AY104236.1  GI:21207314
KEYWORDS    HTC.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 2872)
  AUTHORS   Hainey,C.F., Dolan,M., Miao,G.H., Wallet,J.M., Wallbrandt,M.S.,
            Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
  TITLE     Maize Mapping Project/DuPont Consensus Sequences for Design of
            Overgo Probes
  JOURNAL   Unpublished (2002)
REFERENCE   2  (bases 1 to 2872)
  AUTHORS   Coe,E.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-APR-2002) Maize Mapping Project, University of
            Missouri, Columbia, MO 65211, USA
COMMENT     If you are interested in getting corresponding physical clones,
            these are publicly available from ZmDB and may be found by BLAST
            searching at MSL, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR,
            www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the
            maize cDNA sequences is either Virginia Wallet, Stanford or Pat
            Schnable, Iowa State, then clones may be requested from ZmDB:
            www.zmdb.iastate.edu.
FEATURES             Location/Qualifiers
     source          1. .2872
                     /organism="Zea mays"
                     /mol_type="mRNA"
                     /db_xref="MaizeDB:637084"
                     /db_xref="taxon:4577"
                     /clone_lib="Maize Mapping Project/DuPont Consensus
                     Library"
                     /note="this sequence is part of a project of EST
                     assemblies resulting from the application of public
                     contigs to seed DuPont contigs; this resource was
                     assembled by DuPont as part of a collaboration for the
                     overgo addressing of BACs in conjunction with the Maize
                     Mapping Project"
BASE COUNT      730 a    600 c    688 g    852 t      2 others
ORIGIN
```

Alignment Scores:

Pred. No.:	1.37e-267	Length:	2872
Score:	3149.00	Matches:	592
Percent Similarity:	80.28%	Conservative:	88
Best Local Similarity:	69.89%	Mismatches:	130
Query Match:	60.10%	Indels:	37
DB:	11	Gaps:	10

US-09-980-043-2 (1-978) x AY104236 (1-2872)

```
Qy        151 ProSerAlaGluAlaSerGluProLeuSerIleValTyrProIleProArgAsnLysLeu 170

Db          1 CCACGCGTCCGTGCACGGCTACCTCTAAGTCGCATAGTTCCGATATCTCCAAACGAGCTT 60

Qy        171 ThrProTyrArgAlaValIleIleMetArgLeuValIleLeuGlyLeuPhePheHisPhe 190

Db         61 AACCTTTATCGGATCGTGATTGTTCTCCGGCTTATCATCCTATGTTTCTTCTTTCAATAT 120

Qy        191 ArgIleThrAsnProValAspSerAlaPheGlyLeuTrpLeuThrSerValIleCysGlu 210

Db        121 CGTATAACTCATCCAGTGGAAGATGCTTATGGGTTGTGGCTTGTATCTGTTATTTGTGAA 180

Qy        211 IleTrpPheAlaPheSerTrpValLeuAspGlnPheProLysTrpAsnProValAsnArg 230

Db        181 GTTTGGTTTGCCTTGTCTTGGCTTCTAGATCAGTTCCCAAAGTGGTATCCTATCAACCGT 240

Qy        231 GluThrTyrIleGluArgLeuSerAlaArgTyrGluArgGluGlyGluProSerGlnLeu 250

Db        241 GAAACTTACCTCGATAGACTTGCATTGAGATATGATAGGGAGGGTGAGCCATCCCAGTTG 300

Qy        251 -AlaGlyValAspPhePheValSerThrValAspProLeuLysGluProProLeuIleTh 270

Db        301 GGCTCCAATCGATGTCTTTGTTAGTACAGTGGATCCACTTAAGGAACCTCCTCTAATTAC 360

Qy        270 rAlaAsnThrValLeuSerIleLeuAlaValAspTyrProValAspLysValSerCysTy 290

Db        361 TGGCAACACTGTCCTGTCCATTCTTGCTGTGGATTACCCTGTTGACAAAGTATCATGTTA 420

Qy        290 rValSerAspAspGlyAlaAlaMetLeuSerPheGluSerLeuValGluThrAlaGluPh 310

Db        421 TGTTTCTGATGACGGTTCAGCTATGTTGACTTTTGAAGCGCTATCTGAAACCGCAGAGTT 480

Qy        310 eAlaArgLysTrpValProPheCysLysLysPheSerIleGluProArgAlaProGluPh 330

Db        481 TGCAAGGAAATGGGTTCCCTTTGCAAGAAACACAATATTGAACCTAGGGCTCCAGAGTT 540

Qy        330 eTyrPheSerGlnLysIleAspTyrLeuLysAspLysValGlnProSerPheValLysGl 350

Db        541 TTACTTTGCTCGAAAGATAGATTACCTAAAGGACAAAATACAACCTTCTTTTGTGAAAGA 600

Qy        350 uArgArgAlaMetLysArgAspTyrGluGluTyrLysValArgValAsnAlaLeuValAl 370

Db        601 AAGGCGGGCTATGAAGAGGGAGTGTGAAGAGTTCAAAGTACGGATCGATGCCCTTGTTGC 660

Qy        370 aLysAlaGlnLysThrProGluGluGlyTrpThrMetGlnAspGlyThrProTrpProGl 390

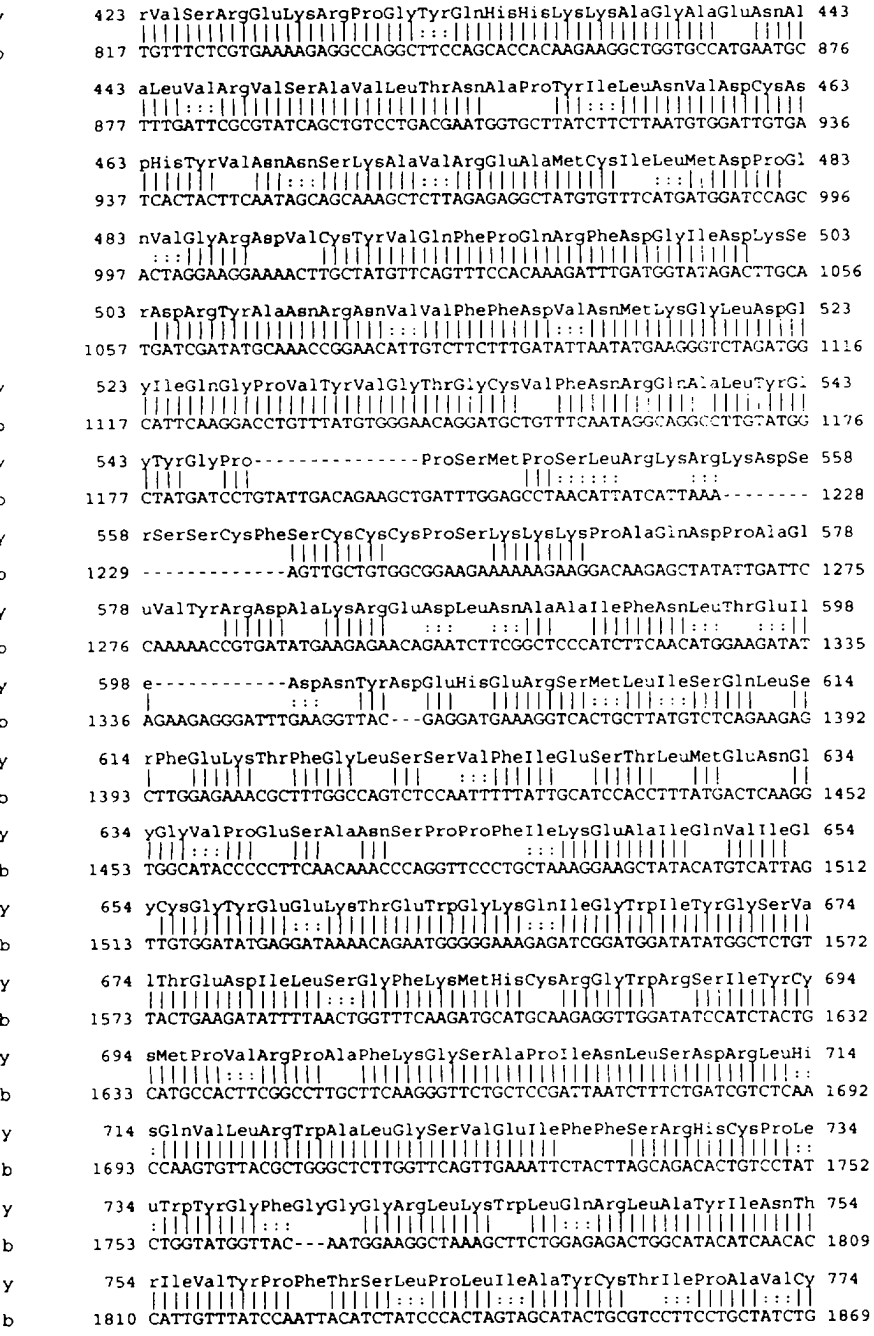
Db        661 AAAAGCGCAAAAAATACCTGAGGAGGGCTGGACCATGGCTGATGGCACTCCYTGGCCTGG 720

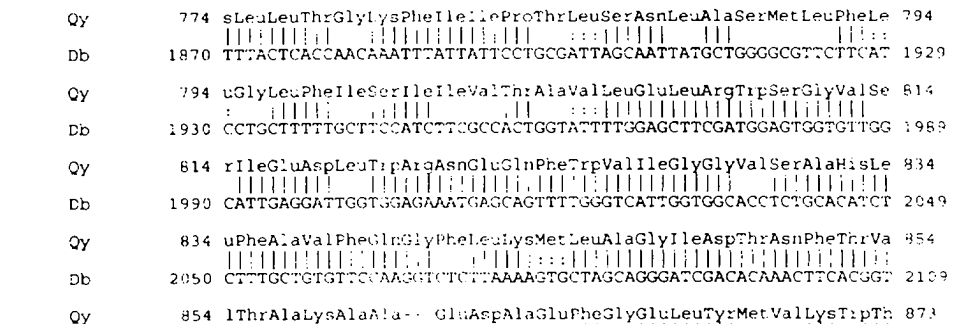
Qy        390 yAsnAsnThrArgAspHisPro--------------------GlyHisAspSerGlyLe 403

Db        721 GAATAACCCTAGAGATCATCCMGGAATGATCCAAGTATTCTTGGGCCACAGTGGTGGGCT 780

Qy        403 uProTrpGluIleLeuGlyAlaArgAspIleGluGlyAsnGluLeuProArgLeuValTy 423

Db        781 T--- -- ----------------GACACGGATGGGAATGAGTTGCCACGGCTTGTTTA 816
```





RESULT 7
AY112236
LOCUS AY112236 3728 bp mRNA linear HTC 17-OCT-2002
DEFINITION Zea mays CL1160 1 mRNA sequence.
ACCESSION AY112236
VERSION AY112236.1 GI:21216826
KEYWORDS HTC.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 3728)
AUTHORS Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S., Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
TITLE Maize Mapping Project/DuPont Consensus Sequences for Design of Overgo Probes
JOURNAL Unpublished (2002)
REFERENCE 2 (bases 1 to 3728)
AUTHORS Coe,E.H.
TITLE Direct Submission
JOURNAL Submitted (25-APR-2002) Maize Mapping Project, University of Missouri, Columbia, MO 65211, USA
COMMENT If you are interested in getting corresponding physical clones, these are publicly available from ZmDB and may be found by BLAST searching at MSL, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR, www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the maize cDNA sequences is either Virginia Walbot, Stanford or Pat Schnable, Iowa State, then clones may be requested from ZmDB:

```
          www.zmdb.iastate.edu.
FEATURES             Location/Qualifiers
     source          1. .3728
                     /organism="Zea mays"
                     /mol_type="mRNA"
                     /db_xref="MaizeDB:630049"
                     /db_xref="taxon:4577"
                     /clone_lib="Maize Mapping Project/DuPont Cornsensus
                     Library"
                     /note="this sequence is part of a project of EST
                     assemblies resulting from the application of public
                     contigs to seed DuPont contigs; this resource was
                     assembled by DuPont as part of a collaboration for the
                     overgo addressing of BACs in conjunction with the Maize
                     Mapping Project"
BASE COUNT      782 a    845 c    925 g    878 t    298 others
ORIGIN

Alignment Scores:
Pred. No.:              1.32e-263       Length:         3728
Score:                  3105.50         Matches:        614
Percent Similarity:     67.81%          Conservative:   119
Best Local Similarity:  56.80%          Mismatches:     218
Query Match:            59.27%          Indels:         132
DB:                     11              Gaps:           18

US-09-980-043-2 (1-978) x AY112236 (1-3728)

Qy          9 CysHisThrCysGlyGluGlnValGlyHisAspAlaAsnGlyGluLeuPheValAlaCys 28
Db        292 TGCCAGATCTGCGGCGATGGCGTGGGCACTACGGCGGAGGCAGACGTCTTCACCGCCTGC 351

Qy         29 HisGluCysSerTyrProMetCysLysSerCysPheGluPheGluIleAsnGluGlyArg 48
Db        352 GACGTCTGCGGGTTCCCGGTGTGCCGCCCCTGCTACGAGTACGAGCGCAAGGACGGCACA 411

Qy         49 LysValCysLeuArgCys---------------------GlySerPro--------- 57
Db        412 CAAGCGTGNNNNNAGTGCNNNNNCAAGTACAAGCGCCACAAGGGGAGTCCAGCGATCCGA 471

Qy         58 ------------------TyrAspGluAsnLeuLeuAspAspValGluLysLysGlySer 71
Db        472 GGGGAGGAAGGAGACGATACTGATGCCGATGATGCTAGCGACTTCAACTACCCTGCATCT 531

Qy         72 GlyAsn--------GlnSerThrMetAlaSerHisLeuA------------------- 82
Db        532 GGCAATGACGACCAGAAGNAGAATTNTGANAGGATGNNAGTTGGGGCATGAATGCT 591

Qy         83 ----------------------------AspSerGluAspValGlyIle ----- 89
Db        592 NNNNNCAGCGGGGATGTTGGCCGCCCCAAGTATGACAGTGGTGAGATCGGGCTTACCAAG 651

Qy         89 ---------------------------- ---------- ---------- 89
Db        652 TACGACAGTGGTGAGATCCCTCGGGGATACATCCCNNNNNNNNNNNNNNNNNATTTCG 711

Qy         89 ---------------------------- ---------- ---------- 89
Db        712 GGAGAAATCCCTGGTGCTTCCCCTGACCATCATATGATGTCTCCTACTGGGAACATTGGC 771

Qy         90 ----------HisAlaArgHisIleSerSerValSerThrValAspSerGlu 103
Db        772 AGGCGCGCCCCATTTCCCTATATGAATCATTCATCAAATCCGTCG---------AGGGAA 822

Qy        104 MetAsnAspGluTyrGlyAsnProIleTrpLysAsnArgValLysSerCysLysAspLys 123
Db        823 TTCTCTGGTAGCGTTGGGAATGTTGCCTGGAAAGAGAGGGTTGATGGCTGGAAAATGAAG 882

Qy        124 GluAsnLys----------LysLysLysArgSerProLysAlaGluThrGluProAla 139
Db        883 CAGGACAAGGGAACAATTCCCATGACGAATGGCACAAGCATTGCTCCCTCTGAGGGCCGG 942

Qy        140 GlnVal----------ProThrGluGlnGlnMetGluGluLys---ProSerAla 153

Db        943 GGTGTTGGTGATATTGATGCATCAACTGATTACAACATGGAAGATGCCTTATTAAACGAT 1002

Qy        154 GluAlaSerGluProLeuSerIleValTyrProIleProArgAsnLysLeuThrProTyr 173
Db       1003 GAAACTCGCCAGCCTCTATCTAGGAAAGTTCCACTTCCTTCCTCCAGGATAAATCCATAC 1062

Qy        174 ArgAlaValIleIleMetArgLeuValIleLeuGlyLeuPhePheHisPheArgIleThr 193
Db       1063 AGGATGGTCATTGTGCTACGATTGATTGTTCTAAGCATCTTCTTGCACTACCGGATCACA 1122

Qy        194 AsnProValAspSerAlaPheGlyLeuTrpLeuThrSerValIleCysGluIleTrpPhe 213
Db       1123 AATCCTGTGCGTAATGCATACCCACTGTGGCTTCTATCTGTTATATGTGAGATCTGGTTT 1182

Qy        214 AlaPheSerTrpValLeuAspGlnPheProLysTrpAsnProValAsnArgGluThrTyr 233
Db       1183 GCTCTTTCCTGGATATTGGATCAGTTTCCAAAGTGGTTTCCAATCAACCGCGAGACTTAC 1242

Qy        234 IleGluArgLeuSerAlaArgTyrGluArgGluGlyGluProSerGlnLeuAlaGlyVal 253
Db       1243 CTTGATAGACTCGCATTAAGGTATGACCGGGAAGGTGAGCCATCTCAGTTGGCTGCTGTT 1302

Qy        254 AspPhePheValSerThrValAspProLeuLysGluProProLeuIleThrAlaAsnThr 273
Db       1303 GACANNNNNGTCAGTACTGTCGACCCAATGAAGGAGCCTCCTCTTGTCACTGCCAATACC 1362

Qy        274 ValLeuSerIleLeuAlaValAspTyrProValAspLysValSerCysTyrValSerAsp 293
Db       1363 GTGCTATCCATTCTCGCTGTGGACTATCCTGTGGATAAGGTCTCTTGCTATGTATCTGAT 1422

Qy        294 AspGlyAlaAlaMetLeuSerPheGluSerLeuValGluThrAlaGluPheAlaArgLys 313
Db       1423 GATGGAGCTGCTATGCTGACATTTGATGCACTAGCTGAGACTTCAGAGTTTGCTAGAAAA 1482

Qy        314 TrpValProPheCysLysLysPheSerIleGluProArgAlaProGluPheTyrPheSer 333
Db       1483 TGGGTGCCATTTGTTAAGAAGTACAACATTGAACCTAGAGCTCCTGAATGGTACTTCTCC 1542

Qy        334 GlnLysIleAspTyrLeuLysAspLysValGlnProSerPheValLysGluArgArgAla 353
Db       1543 CAGAAAATTGATTACTTGAAGGACAAAGTGCACCCTTCATTTGTTAAAGACCGCCGGGNN 1602

Qy        354 MetLysArgAspTyrGluGluTyrLysValArgValAsnAlaLeuValAlaLysAlaGln 373
Db       1603 NNNNNNNNNNNNNNNNNNNNNCAAAATTAGGGTAAATGGCCTTGTTGCTAAGGCACNN 1662

Qy        374 LysThrProGluGluGlyTrpThrMetGlnAspGlyThrProTrpProGlyAsnAsnThr 393
Db       1663 NNNGTCCCTGAGGAAGGATGGATCATGCAAGATGGCACACCATGGCCAGGAAACAATACC 1722

Qy        394 ArgAspHisProGlyHisAspSerGlyLeuProTrpGluIleLeuGlyAlaArgAspIle 413
Db       1723 AGGGACCATCCTGGAAT-GATTCAGGTTTTCCTTGGTCAC--AGTGGTGGTCTTGATACT 1779

Qy        414 GluGlyAsnGluLeuProArgLeuValTyrValSerArgGluLysArgProGlyTyrGln 433
Db       1780 GAGGGTAATGAGCTACCCCGTTTGGTCTATGTTTCTCGTGNNNNNCGTCCTGGATTCCAG 1839

Qy        434 HisHisLysLysAlaGlyAlaGluAsnAlaLeuValArgValSerAlaValLeuThrAsn 453
Db       1840 CATCACAAGAAAGCTGGTGCCATGAATGCTCTTGTCCGCGTCTCAGCTGTGCTTACCAAT 1899

Qy        454 AlaProTyrIleLeuAsnValAspCysAspHisTyrValAsnAsnSerLysAlaValArg 473
Db       1900 GGACAATACATGTTGAATCTTGATTGTGATCACTACATCAACAACAGTAAGGCTCTCAGG 1959

Qy        474 GluAlaMetCysIleLeuMetAspProGlnValGlyArgAspValCysTyrValGlnPhe 493
Db       1960 GAAGCTATGTGCTTCCTTATGGATCCTAACCTAGGAAGGAGTGTCTGCTATGTTCAGTTT 2019

Qy        494 ProGlnArgPheAspGlyIleAspLysSerAspArgTyrAlaAsnArgAsnValValPhe 513
```

```
2020 CCCCAGAGGTTCGATGGTATTGATAGGAATGATCGATATGCCAACAGGAACACCGTGNNN 2079

 514 PheAspValAsnMetLysGlyLeuAspGlyIleGlnGlyProValTyrValGlyThrGly 533
     |||:::|||:::::|||||||||||||||||||||||||||||||||||||||||||||
2080 NNCGATATTAACTTGAGAGGTCTTGATGGCATCCAAGGACCAGTTTATGTGGGCACTGGC 2139

 534 CysValPheAsnArgGlnAlaLeuTyrGlyTyrGlyProProSerMetProSerLeuArg 553
     ||||||||||||||||   ||||||||||||||                      ::::::
2140 TGTGTTTTCAACAGAACAGCTCTATATGGTTAT-----------GAGNNNNNAATTAAG 2187

 554 LysArgLysAsp---------SerSerSerCysPheSerCysCysCysProSerLysLys 570
     ::::::|||              ||||||   |||
2188 CAAAAGAAGGGTGGTTTCTTGTCATCACTATGTGGNNNNNNNNNNNNNNNNNNNNNNNCA 2247

 571 LysProAlaGlnAspPro-------------------------------AlaGluVal 579
     |||         |||
2248 AAGAAGGGCTCAGACAAGNNNNNGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2307

 580 TyrArgAspAlaLysArgGluAspLeuAsnAlaAlaIlePheAsnLeuThrGluIleAsp 599
                 |||   :::       |||   |||
2308 NNNNNNNNNNNNTAGAGGAGGGAGTTGAAGGCGCTGGATTT----------------- 2349

 600 AsnTyrAspGluHisGluArgSerMetLeuIleSerGlnLeuSerPheGluLysThrPhe 619
           :::   |||:::|||:::|||:::||||||:::|||   |||||   |||
2350 ---------GATGATGAGAAATCACTTCTTATGTCTCAAATGAGCTTGGAGAAGAGATTT 2400

 620 GlyLeuSerSerValPheIleGluSerThrLeuMetGluAsnGlyGlyValProGluSer 639
     |||   |||:::   |||:::   ||||||||||||||||
2401 GGCCAATCTGCAGCTTTTGTTGCGTCCACTCTGATGGAATNNNNNNNNNNNNNNNNNNNN 2460

 640 AlaAsnSerProProPheIleLysGluAlaIleGlnValIleGlyCysGlyTyrGluGlu 659
                                 |||   ||||||   ||||||||||||:::
2461 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNTATCCATGTCATAAGTTGTGGCTACGAGGAC 2520

 660 LysThrGluTrpGlyLysGlnIleGlyTrpIleTyrGlySerValThrGluAspIleLeu 679
     |||   |||||||||   :::||||||||||||||||||||||||||||||||||||||
2521 AAGATTGAATGGGGAACTGAGATTGGGTGGATCTATGGTTCTGTGACGGAAGATATTCTC 2580

 680 SerGlyPheLysMetHisCysArgGlyTrpArgSerIleTyrCysMetProValArgPro 699
     :::|||||||||||||||   ||||||||||||||||||||||||||||   ||||||
2581 ACTGGGTTCAAGATGCACGCACGAGGCTGGCGGTCGATCTACTGCATGCCTAAGCGGCCG 2640

 700 AlaPheLysGlySerAlaProIleAsnLeuSerAspArgLeuHisGlnValLeuArgTrp 719
     ||||||||||||||||||||||||||||||||||||||:::|||||||||||||||||
2641 GCCTTCAAGGGATCGGCTCCCATCAATCTCTCAGACCGTCTGAACCAGGTGCTCCGGTGG 2700

 720 AlaLeuGlySerValGluIlePhePheSerArgHisCysProLeuTrpTyrGlyPheGly 739
     ||||||||||||||||||   ||||||||||         |||||||:::
2701 GCTCTCGGTTCAGTGGAAATCCTTTTCAGCCGGCATTGNNNNNTATGGTACGGGTAC--- 2757

 740 GlyGlyArgLeuLysTrpLeuGlnArgLeuAlaTyrIleAsnThrIleValTyrProPhe 759
     ||||||||||||||:::|||:::|||   ||||||||||||||   :::||||||
2758 GGAGGACGCCTGAAGTTCTTGGAGAGATTCGCCTACATCAACACCACCATCTACCCGCTC 2817

 760 ThrSerLeuProLeuIleAlaTyrCysThrIleProAlaValCysLeuLeuThrGlyLys 779
     ||||||||||||||:::   ||||||   :::|||||||:::||||||||||||||||||
2818 ACGTCCCTCCCGCTCCTCATTTACTGTATCCTGCCTGCCATCTGCCTGCTCACGGGGAAG 2877

 780 PheIleIleProThrLeuSerAsnLeuAlaSerMetLeuPheLeuGlyLeuPheIleSer 799
     ||||||||||||   :::||||||   ||||||:::   |||:::   ||||||||||||
2878 TTCATCATCCCAGAGATCAGCAACTTCGCTAGTATCTGGTTCATCTCTCTCTTCATCTCG 2937

 800 IleIleValThrAlaValLeuGluLeuArgTrpSerGlyValSerIleGluAspLeuTrp 819
     |||       |||   :::||||||:::||||||||||||||||   |||::::::   |||
2938 ATCTTCGCCACGGGTATCCTGGAGATGAGGTGGAGCGGCGTGGGCATCGACGAGTGGTGG 2997

 820 ArgAsnGluGlnPheTrpValIleGlyGlyValSerAlaHisLeuPheAlaValPheGln 839
     ||||||||||||||||||||||||||||:::||||||||||||||||||||||||||||
2998 AGGAACGAGCAGTTCTGGGTCATCGGAGGCATCTCCGCCCACCTCTTCGCCGTCTTCCAG 3057

 840 GlyPheLeuLysMetLeuAlaGlyIleAspThrAsnPheThrValThrAlaLysAlaAla 859
     |||   |||||:::||||||||||||||||||||||||||||||:::|||||:::
3058 GGCCTCCTCAAGGTGCTTGCCGGCATCGACACCAACTTCACCGTCACCTCCAAGGCCTCG 3117
```

```
Qy        860 ---GluAspAlaGluPheGlyGluLeuTyrMetValLysTrpThrThrLeuLeuIlePro 878
              ||||||   :::|||   |||||:|||||   |||||||||||||||||||||||
Db       3118 GATGAAGACGGCGACTTCGCGGAGCTGTACATGTTCAAGTGGACGACACTTCTGATCCCG 3177

Qy        879 ProThrThrLeuLeuIleIleAsnMetSerGly---CysAlaGlyPheSerAspAlaLeu 897
              |||||||||:::||:|||||||.||:::   |.|       ||||||   |||   |||:::
Db       3178 CCCACCACCATCCTCATCATCAACCTGGTCGGCGTTGTTGCCGGCATCTCCTACGCCATC 3237

Qy        898 AsnLysGlyTyrGluAlaTrpGlyProLeuPheGlyLysValPhePheAlaPheTrpVal 917
              |||   |||||:::::|||||||.|||||:|||||:|:::|||||||||||||||||.||
Db       3238 AACAGCGGGTACCAGTCGTGGGGTCCGCTCTTCGGCAAGCTCTTCTTCGCCTTCTGGGTG 3297

Qy        918 IleLeuHisLeuTyrProPheLeuLysGlyLeuMetGlyArgGlnAsnLeuThrProThr 937
              |||:::|||||||||::||||||||||||||||||||||||||||||   ||||||||||
Db       3298 ATCGTTCACCTGTACCCGTTCCTCAAGGGTCTCATGGGTCGGCAGAACCGCACCCCGACC 3357

Qy        938 IleValValLeuTrpSerValLeuLeuAlaSerValPheSerLeuValTrpValLysIle 957
              ||||:|||||:::||  ::::::||'|||||:'||:::|||||||||:::|||||:::|||
Db       3358 ATCGTGGTTGTCTGGTCGGATCCTGCTTGTCGTCTATCTTCTCCTTGCTGTGGGTTCGCATC 3417

Qy        958 AsnProPheValAsnLysValAspAspThrLeuValAlaGluThrCysIleSerIleAsp 977
              :::||||   ::  :::::|                        |||||      ||:::
Db       3418 GATCCGTTCACCAA'TCGTCAC'TGGCCCG-  GATACTCGAACGTGT- -GGCATCAAC 3471

Qy        978 Cys 978
              |||
Db       3472 TGC 3474

RESULT 8
AY110079
LOCUS       AY110079                3696 bp    mRNA    linear   HTC 17-OCT-2002
DEFINITION  Zea mays CL1164_1 mRNA sequence.
ACCESSION   AY110079
VERSION     AY110079.1  GI:21214162
KEYWORDS    HTC.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 3696)
  AUTHORS   Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S.,
            Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
  TITLE     Maize Mapping Project/DuPont Consensus Sequences for Design of
            Overgo Probes
  JOURNAL   Unpublished (2002)
REFERENCE   2  (bases 1 to 3696)
  AUTHORS   Coe,E.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-APR-2002) Maize Mapping Project, University of
            Missouri, Columbia, MO 65211, USA
COMMENT     If you are interested in getting corresponding physical clones,
            these are publicly available from ZmDB and may be found by BLAST
            searching at MSU, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR,
            www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the
            maize cDNA sequences is either Virginia Walbot, Stanford or Pat
            Schnable, Iowa State, then clones may be requested from ZmDB:
            www.zmdb.iastate.edu.
FEATURES             Location/Qualifiers
     source          1..3696
                     /organism="Zea mays"
                     /mol_type="mRNA"
                     /db_xref="MaizeDB:630055"
                     /db_xref="taxon:4577"
                     /clone_lib="Maize Mapping Project/DuPont Cornsensus
                     Library"
                     /note="this sequence is part of a project of EST
                     assemblies resulting from the application of public
                     contigs to seed DuPont contigs; this resource was
                     assembled by DuPont as part of a collaboration for the
                     overgo addressing of BACs in conjunction with the Maize
                     Mapping Project"
```

```
BASE COUNT      831 a    766 c    880 g    945 t    274 others
ORIGIN

Alignment Scores:
Pred. No.:              3.01e-260       Length:         3696
Score:                  3067.50         Matches:        599
Percent Similarity:     68.06%          Conservative:   136
Best Local Similarity:  55.46%          Mismatches:     216
Query Match:            58.54%          Indels:         129
DB:                     11              Gaps:           19

US-09-980-043-2 (1-978) x AY110079 (1-3696)

Qy          8 IleCysHisThrCysGlyGluGlnValGlyHisAspAlaAsnGlyGluLeuPheValAla 27
Db        302 GTGTGCCAGATCTGCGGCGACGGCGTGGGCACCGCGGCGGACGGCGACCTCTTCACCGCC 361

Qy         28 CysHisGluCysSerTyrProMetCysLysSerCysPheGluPheGluIleAsnGluGly 47
Db        362 TGCGACGTCTGCGGCTTCCCCGTGTGCCGCCCATGCTACGAGTACGAGCGCAAGGACGGC 421

Qy         48 ArgLysValCysLeuArgCys--- ------------------GlySerPro------ 57
Db        422 ACCCAGGCGTGCCCGCAGTGCAAGACTAAGTACAAGCGCCACAAAGGGAGCCCACCAGTA 481

Qy         58 ------------------ -TyrAspGluAsnLeuLeuAspAspValGluLysLysGly 70
Db        482 CACGGTGAGGAAAATGAGGATGTGGATGCTGACGATGTGAGTGACTACAACTACCAAGCA 541

Qy         71 SerGlyAsnGln---------- ---------- ---------- 74
Db        542 TCTGGCAACCAGGATCAGAAGCAAAAGATTGCTGAGAGAATGCTCACTTGGCGGACAAAC 601

Qy         75 - - - -SerThrMetAlaSerHisLeuAsnAspSerGlnAspValGlyIleHisAla 91
Db        602 TCACGTGGCAGTGATATTGGCCTGGCTAAGTATGACAGCGGTGAAATTGGG---CATGGG 658

Qy         92 ArgHis--- --------- ---- ---------- - - ---------- 93
Db        659 AAGTATGACAGTGGTGAGATCCCTGGTGGATATATCCCGTCACTAACTCATAGCCAGATC 718

Qy         94 ---------- ------ - - IleSerServalSerThrVal--- 100
Db        719 TCAGGAGAGATTCCTGGAGCTTCCCCTGATCATATGATGTCTCCTGTTGGGAACATTGGC 778

Qy        101 -- ---------- - AspSerGluMetAsn 105
Db        779 AGGCGTGGACATCAATTTCCTTATGTAAATCATTCTCCAAACTATCGAGGAGTTCTCC 838

Qy        106 AspGluTyrGlyAsnProIleTrpLysAsnArgValLysSerCysLysAspLysGluAsn 125
Db        839 GGTAGCCTTGGCAATGTTGCATGGAAAGAGAGGGTGGATGGATGGAAAATGAAGGATAAA 898

Qy        126 LysLysLysLysArgSerProLysAlaGluThrGluProAlaGln ---------- 140
Db        899 GGTGCAATTCCTATGACCAATGGAACAAGCATTGCTCCATCAGAAGGGCGTGGAGTTGCT 958

Qy        141 ---------ValProThrGluGlnGlnMetGluGluLys-- ProSerAlaGluAlaSer 156
Db        959 GATATTGATGCTTCTACTGATTATAACATGGAAGATGCCTTACTGAATGATGAAACTCGG 1018

Qy        157 GluProLeuSerIleValTyrProIleProArgAsnLysLeuThrProTyrArgAlaVal 176
Db       1019 CAACCTCTATCTAGAAAAGTGCCAATTCCTTCATCCAGAATAAATCCGTACAGAATGGTC 1078

Qy        177 IleIleMetArgLeuValIleLeuGlyLeuPhePheHisPheArgIleThrAsnProVal 196
Db       1079 ATTGTGCTACGTTTGGCTGTTCTATGCATATTCTTGGCGCTACCGTATCACACATCCTGTG 1138

Qy        197 AspSerAlaPheGlyLeuTrpLeuThrSerValIleCysGluIleTrpPheAlaPheSer 216
Db       1139 AACAATGCATATCCACTGTGGCTTTTATCCGTCATATGTGAGATCTGGTTTGCTTTGTCC 1198

Qy        217 TrpValLeuAspGlnPheProLysTrpAsnProValAsnArgGluThrTyrIleGluArg 236
Db       1199 TGGATTTTGGATCAGTTCCCAAAGTGGTCCCCAATCAACCGTGAAACATACCTTGATAGA 1258

Qy        237 LeuSerAlaArgTyrGluArgGluGlyGluProSerGlnLeuAlaGlyValAspPhePhe 256
Db       1259 CTGGCTTTAAGGTATGACCGAGAAGGTGAACCATCTCAATTAGCTCCTGTTGATANNNNN 1318

Qy        257 ValSerThrValAspProLeuLysGluProProLeuIleThrAlaAsnThrValLeuSer 276
Db       1319 GTCAGTACTGTGGATCCAATGAAGGAGCCTCCTCTTGTCACTGCAAATACTGTGCTTTCC 1378

Qy        277 IleLeuAlaValAspTyrProValAspLysValSerCysTyrValSerAspAspGlyAla 296
Db       1379 ATCCTTGCTGTCGATTATCCGGTTGACAAGGTATCTTGCTATGTTTCGGATGATGGAGCT 1438

Qy        297 AlaMetLeuSerPheGluSerLeuValGluThrAlaGluPheAlaArgLysTrpValPro 316
Db       1439 GCTATGCTGACTTTTGATGCTCTCTCTGAAACTTCAGAGTTTGCTAGAAAATGGGTTCCG 1498

Qy        317 PheCysLysLysPheSerIleGluProArgAlaProGluPheTyrPheSerGlnLysIle 336
Db       1499 TTCTGTAAGAAGTACAACATAGAGCCTANGGCCCCGGAATGGTACTTTGCTCAGAAAATT 1558

Qy        337 AspTyrLeuLysAspLysValGlnProSerPheValLysGluArgArgAlaMetLysArg 356
Db       1559 GATTACTTGAAAGACAAAGTTCAAACCTCATTTGTGAAAGAACGCCGGGNNNNNNNNNNN 1618

Qy        357 AspTyrGluGluTyrLysValArgValAsnAlaLeuValAlaLysAlaGlnLysThrPro 376
Db       1619 NNNNNNNNNNNNNCAAAGTTCGTATCAATGGTCTTGTAGCCAAGGCACNNNNNGTTCCC 1678

Qy        377 GluGluGlyTrpThrMetGlnAspGlyThrProTrpProGlyAsnAsnThrArgAspHis 396
Db       1679 GAGGAGGGATGGATCATGCAAGATGGTACACCTTGGCCTGGGAACAATACTAGGGACCAT 1738

Qy        397 Pro--------------------GlyHisAspSerGlyLeuProTrpGluIleLeuGly 409
Db       1739 CCTGGAATGATTCAGGTTTTCCTGGGTCACAGTGGAGGGCTT------------------ 1780

Qy        410 AlaArgAspIleGluGlyAsnGluLeuProArgLeuValTyrValSerArgGluLysArg 429
Db       1781 ------GACGTTGAAGGCAATGAACTTCCTCGTTTGGTTTATGTGTCTCGTGNNNNNCGT 1834

Qy        430 ProGlyTyrGlnHisHisLysLysAlaGlyAlaGluAsnAlaLeuValArgValSerAla 449
Db       1835 CCTGGATTCCAACATCACAAGAAGGCTGGTGCCATGAATGCACTTGTTCGTGTATCAGCT 1894

Qy        450 ValLeuThrAsnAlaProTyrIleLeuAsnValAspCysAspHisTyrValAsnAsnSer 469
Db       1895 GTCCTTACTAATGGGCAATACATGTTGAATCTTGATTGTGACCACTACATCAATAATAGC 1954

Qy        470 LysAlaValArgGluAlaMetCysIleLeuMetAspProGlnValGlyArgAspValCys 489
Db       1955 AAGGCTCTTCGAGAAGCTATGTGCTTCCTTATGGACCCAAACCTAGGAAGGAATGTCTGT 2014

Qy        490 TyrValGlnPheProGlnArgPheAspGlyIleAspLysSerAspArgTyrAlaAsnArg 509
Db       2015 TATGTCCAATTTCCTCAGAGGTTTGATGGTATTGATAGGAATGACCGATATGCAAACAGG 2074

Qy        510 AsnValValPhePheAspValAsnMetLysGlyLeuAspGlyIleGlnGlyProValTyr 529
Db       2075 AACACTGTGNNNNNCGATATTAACTTGAGAGGTCTTGACGGCATTCAAGGGCCAGTTTAT 2134

Qy        530 ValGlyThrGlyCysValPheAsnArgGlnAlaLeuTyrGlyTyrGlyProPro------ 547
Db       2135 GTGGGAACTGGTTGTGTGTTTAACAGAACGGCCTTATATGGTTATGAGCCTCCAGTCAAG 2194

Qy        548 ---SerMetProSerLeuArgLysArgLysAspSerSerSerCysPheSerCysCysCys 566
Db       2195 NNNNNNNGCCAGGCTTC------------TTCTCTTCGCTTTGTNNNNNAAGGNNNNG 2242

Qy        567 ProSerLysLysLysProAlaGlnAspProAlaGluValTyrArgAspAlaLysArgGlu 586
```

```
o       2243 ACGTCAAAATCTAAGAAGAGCTCGGAAAAGAAGAAGTCACATAGACACGCA--------- 2293

y        587 AspLeuAsnAlaAlaIlePheAsnLeuThrGluIleAsp----------------Asn 600

b       2294 GACAGTTCTGTACCAGTATTTAATCTCGAAGATATAGAGGAAGGGATTGAAGGTTCTCAG 2353

y        601 TyrAspGluHisGluArgSerMetLeuIleSerGlnLeuSerPheGluLysThrPheGly 620

b       2354 TTTGATGAT---GAGAAATCGCTGATTATGTCTCAAATGAGCTTGGAGAAGAGATTTGGC 2410

y        621 LeuSerSerValPheIleGluSerThrLeuMetGluAsnGlyGlyValProGluSerAla 640

o       2411 CAGTCCAGTGNNNNNGTAGCCTCTACTCTGATGGAATATGGTGGTGTTCCACAATCTGCA 2470

y        641 AsnSerProProPheIleLysGluAlaIleGlnValIleGlyCysGlyTyrGluGluLys 660

o       2471 ACTCCNNNNNNNNNNNNNNNNNNNNTATTCATGTCATCAGCTGTGGCTATGAGGACAAA 2530

y        661 ThrGluTrpGlyLysGlnIleGlyTrpIleTyrGlySerValThrGluAspIleLeuSer 680

b       2531 ACTGACTGGGGAACTGAGATTGGGTGGATCTATGGTTCTGTTACAGAAGACATTCTCACC 2590

y        681 GlyPheLysMetHisCysArgGlyTrpArgSerIleTyrCysMetProValArgProAla 700

o       2591 GGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCTAAGCGACCAGCT 2650

y        701 PheLysGlySerAlaProIleAsnLeuSerAspArgLeuHisGlnValLeuArgTrpAla 720

b       2651 TTCAAGGGATCTGCTCCTATCAACCTTTCGGATCGTTTGAATCAAGTGCTTCGGTGGGCT 2710

y        721 LeuGlySerValGluIlePhePheSerArgHisCysProLeuTrpTyrGlyPheGlyGly 740

b       2711 CTTGGTTCCATTGAAATTCTTTTCAGCAGGCATTGTCCCATATGGTATGGCTAT---GGA 2767

y        741 GlyArgLeuLysTrpLeuGlnArgLeuAlaTyrIleAsnThrIleValTyrProPheThr 760

b       2768 GGCCGGCTTAAATTCCTGGAGAGATTTGCTTATATCAACACAACAATTTATCCACTCACA 2827

y        761 SerLeuProLeuIleAlaTyrCysThrIleProAlaValCysLeuLeuThrGlyLysPhe 780

b       2828 TCAATCCCGCTCCTCCTGTACTGCATATTGCCAGCAGTTTGTCTTCTCACTGGGAAGTTC 2887

y        781 IleIleProThrLeuSerAsnLeuAlaSerMetLeuPheLeuGlyLeuPheIleSerIle 800

b       2888 ATCATCCCAAAGATTAGTAACCTAGAGAGTGTTTGGTTTATATCGCTCTTTATCTCAATC 2947

y        801 IleValThrAlaValLeuGluLeuArgTrpSerGlyValSerIleGluAspLeuTrpArg 820

b       2948 TTTGCCACTGGTATCCTTGNNNNNNNNNNNNNNNNNNNNNNCATTGATGAATGGTGGAGG 3007

y        821 AsnGluGlnPheTrpValIleGlyGlyValSerAlaHisLeuPheAlaValPheGlnGly 840

b       3008 AACGAGCAGTTCTGGGTCATTGGTGGTATTTCTGCGCATTTATTTGCCGTCTTCCAGGGT 3067

y        841 PheLeuLysMetLeuAlaGlyIleAspThrAsnPheThrValThrAlaLysAlaAla--- 859

o       3068 CTCCTGAAGGTGCTTGCTGGTATCGACACGAGCTTCACTGTCACCTCTAAGGCCACTGAC 3127

y        860 GluAspAlaGluPheGlyGluLeuTyrMetValLysTrpThrThrLeuLeuIleProPro 879

b       3128 GAAGAAGGTGATTTTGCCGAGCTCTACATGTTCAAGTGGACAACGCTTCTGATCCCACCA 3187

y        880 ThrThrLeuLeuIleIleAsnMetSerGly---CysAlaGlyPheSerAspAlaLeuAsn 898

b       3188 ACCACTATTTTGATCATCAACCTGGTCGGCGTGGTCGCTGGCATTTCCTACGCAATCAAT 3247

y        899 LysGlyTyrGluAlaTrpGlyProLeuPheGlyLysValPhePheAlaPheTrpValIle 918

b       3248 AGCGGTTACCAGTCATGGGGACCTCTTTTCGGGAAGCTCTTCTTTGCGTTCTGGGTGATT 3307

y        919 LeuHisLeuTyrProPheLeuLysGlyLeuMetGlyArgGlnAsnLeuThrProThrIle 938

Db      3308 GTCCACCTGTACCNTTTCCTCAAGGGCTCATGGGGAAGCAGAACCGCACGCCGACCATT 3367

Qy       939 ValValLeuTrpSerValLeuLeuAlaSerValPheSerLeuValTrpValLysIleAsn 958

Db      3368 GTCGTTGTGTGGNTATGCTGCTTGGGTCGATCTTTTCCCTGATGTGGGTTCGTATCGAT 3427

Qy       959 ProPheValAsnLysValAspAsnThrLeuValAlaGluThrCysIleSerIleAspCys 978

Db      3428 CCATTCACCACCNNNTTCACTGGTCCTGATATCGCGAAA------TGTGGCATCAACTGC 3481

RESULT 9
AY107656
LOCUS       AY107656                1189 bp    mRNA    linear   HTC 16-OCT 2002
DEFINITION  Zea mays PCO126464 mRNA sequence.
ACCESSION   AY107656
VERSION     AY107656.1  GI:21210734
KEYWORDS    HTC.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 1189)
  AUTHORS   Haney,C.F., Lee,M., Wang,C.H., Vogel,J.M., Whitsitt,M.S.,
            Arthur,L.W., Hussaley,M., Morgante,M. and Tingey,S.V.
  TITLE     Maize Mapping Project/DuPont Consensus Sequences for Design of
            Overgo Probes
  JOURNAL   Unpublished (2002)
REFERENCE   2  (bases 1 to 1189)
  AUTHORS   Coe,E.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-APR-2002) Maize Mapping Project, University of
            Missouri, Columbia, MO 65211, USA
COMMENT     If you are interested in getting corresponding physical clones,
            these are publicly available from ZmDB and may be found by BLAST
            searching at MSU, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR,
            www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the
            maize cDNA sequences is either Virginia Walbot, Stanford or Pat
            Schnable, Iowa State, then clones may be requested from ZmDB:
            www.zmdb.iastate.edu.
FEATURES             Location/Qualifiers
     source          1..1189
                     /organism="Zea mays"
                     /mol_type="mRNA"
                     /db_xref="MaizeDB:637366"
                     /db_xref="taxon:4577"
                     /clone_lib="Maize Mapping Project/DuPont Cornsensus
                     Library"
                     /note="this sequence is part of a project of EST
                     assemblies resulting from the application of public
                     contigs to seed DuPont contigs; this resource was
                     assembled by DuPont as part of a collaboration for the
                     overgo addressing of BACs in conjunction with the Maize
                     Mapping Project"
BASE COUNT      215 a    374 c    331 g    269 t
ORIGIN

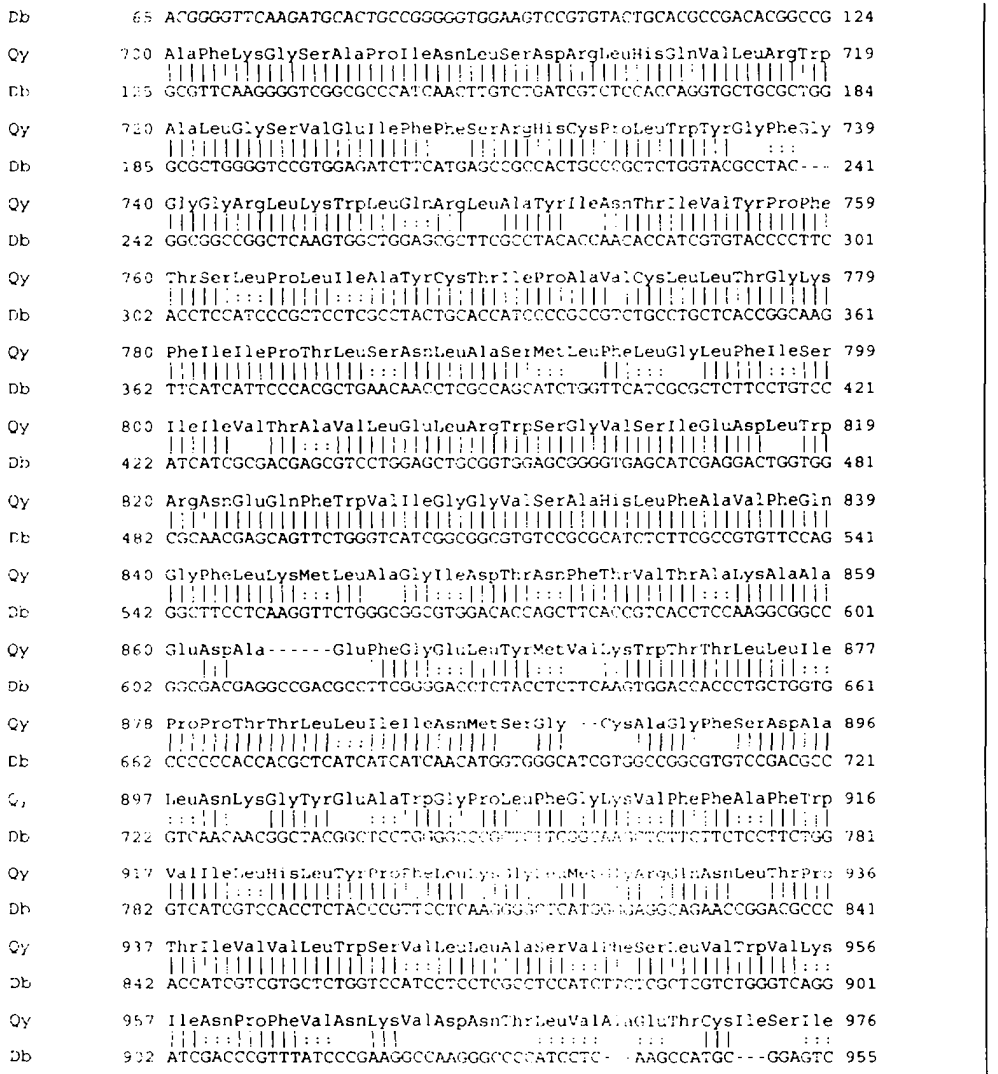
Alignment Scores:
Pred. No.:              2.56e-111       Length:         1189
Score:                  1378.00         Matches:        251
Percent Similarity:     89.44%          Conservative:   37
Best Local Similarity:  77.95%          Mismatches:     28
Query Match:            26.42%          Indels:         6
DB:                     11              Gaps:           5

US-09-980-043-2 (1-978) x AY107656 (1-1189)

Qy       660 LysThrGluTrpGlyLysGlnIleGlyTrpIleTyrGlySerValThrGluAspIleLeu 679

Db         5 GAGACCGAGTGGGGCAAGGAGATTGGGTGGATCTATGGGTCGGTGACAGAGGATATCCTG 64

Qy       680 SerGlyPheLysMetHisCysArgGlyTrpArgSerIleTyrCysMetProValArgPro 699
```

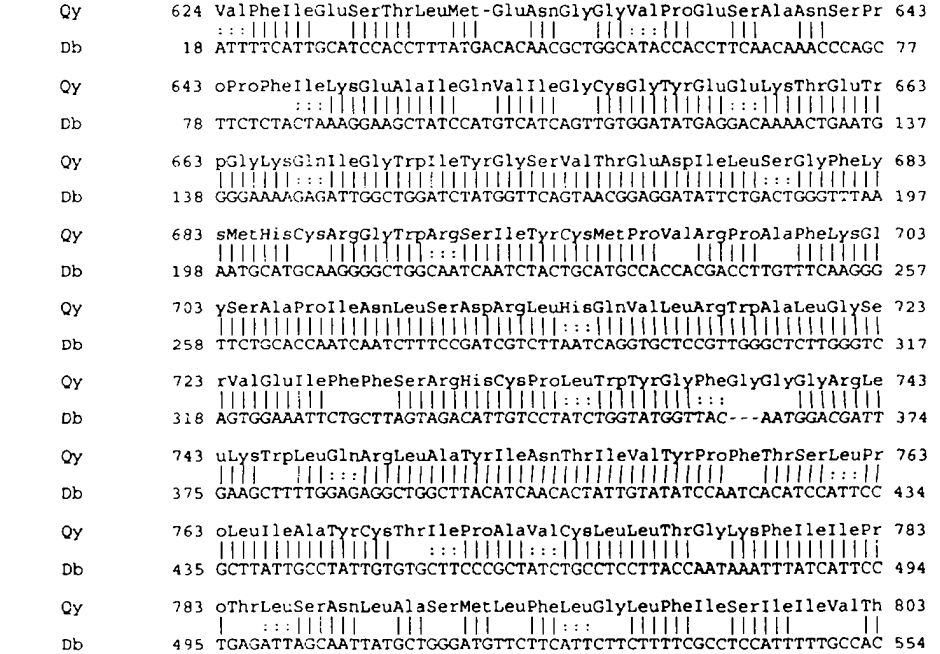


RESULT 10
CD439980
LOCUS CD439980 953 bp mRNA linear EST 03-JUN-2003
DEFINITION EL01N0531D05.b Endosperm_5 Zea mays cDNA, mRNA sequence.
ACCESSION CD439980
VERSION CD439980.1 GI:31355623
KEYWORDS EST.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 953)
AUTHORS Lai,J., Dey,N., Kim,C.S., Becraft,P., Larkins,B., Linton,E. and Messing,J.
TITLE Sequencing of the maize endosperm ESTs
JOURNAL Unpublished
COMMENT Contact: Lai, Jinsheng
Dr. Joachim Messing's lab
Waksman Institute, Rutgers University
190 Frelinghuysen Rd., Piscataway, NJ 08854, USA
Tel: 732-445-3801
Fax: 732-445-5735
Email: jlai@waksman.rutgers.edu
Seq primer: T3.
FEATURES Location/Qualifiers
source 1. .953
/organism="Zea mays"
/mol_type="mRNA"
/cultivar="W22"
/db_xref="taxon:4577"
/tissue_type="Endosperm of 7-23DAP"
/clone_lib="Endosperm_5"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT 217 a 220 c 222 g 294 t
ORIGIN

Alignment Scores:

Pred. No.:	2.66e-101	Length:	953
Score:	1263.50	Matches:	233
Percent Similarity:	84.35%	Conservative:	31
Best Local Similarity:	74.44%	Mismatches:	45
Query Match:	24.11%	Indels:	4
DB:	14	Gaps:	3

US-09-980-043-2 (1-978) x CD439980 (1-953)



```
803 rAlaValLeuGluLeuArgTrpSerGlyValSerIleGluAspLeuTrpArgAsnGluGl 823
555 TGGTATATTGGAGCTTAGATGGAGCGGTGTTGGCATTGAAGATTGGTGGAGAAATGAGCA 614

823 nPheTrpValIleGlyGlyValSerAlaHisLeuPheAlaValPheGlnGlyPheLeuLy 843
615 GTTTTGGGTTATTGGTGGCACCTCTGCCCATCTCTTCGCAGTGTTCCAGGGTCTGCTGAA 674

843 sMetLeuAlaGlyIleAspThrAsnPheThrValThrAlaLysAlaAla---GluAspAl 862
675 AGTGTTGGCTGGGATTGATACCAACTTCACAGTTACCTCAAAGGCATCTGATGAGGATGG 734

862 aGluPheGlyGluLeuTyrMetValLysTrpThrThrLeuLeuIleProProThrThrLe 882
735 CGACTTTGCTGAGCTATATGTGTTCAAGTGGACCAGTTTGCTCATTCCTCCGACCACTGT 794

882 uLeuIleIleAsnMetSerGly---CysAlaGlyPheSerAspAlaLeuAsnLysGlyTy 901
795 TCTTGTCATTAACCTGGTCGGAATGGTGGCAGGAATTTCGTATGCCATTAACAGTGGCTA 854

901 rGluAlaTrpGlyProLeuPheGlyLysValPhePheAlaPheTrpValIleLeuHisLe 921
855 CCAATCCTGGGGTCCGCTCTTTGGAAAGCTGTTCTTCTCGATCTGGGTGATCCTCCATCT 914

921 uTyrProPheLeuLysGlyLeuMetGlyArgGlnAsn 933
915 CTACCCCCTTCTCAAGGGTCTCATGGGAAGGCAAAAC 951
```

```
SULT 11
049602
CUS       AB049602                960 bp    mRNA    linear   EST 10-APR-2002
FINITION  AB049602 Prunus persica fruit Prunus persica cDNA clone pfPpCesA1
          similar to cellulose synthase, mRNA sequence.
CESSION   AB049602
RSION     AB049602.1  GI:20135516
YWORDS    EST.
URCE      Prunus persica (peach)
ORGANISM  Prunus persica
          Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
          Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
          ; eurosids I; Rosales; Rosaceae; Amygdaloideae; Prunus.
FERENCE   1  (bases 1 to 960)
AUTHORS   Hayama,H.
TITLE     Changes in the levels of mRNAs for putative cell wall-related genes
          during peach fruit development
JOURNAL   Unpublished
MMENT     Contact: Hiroko Hayama
          Department of Pomology
          National Institute of Fruit Tree Science
          2-1 Fujimoto, Tsukuba, Ibaraki 305-8605, Japan
          Tel: 81-298-38-6502
          Fax: 81-298-38-6437
          Email: hhiroko@fruit.affrc.go.jp.
ATURES              Location/Qualifiers
   source           1. .960
                    /organism="Prunus persica"
                    /mol_type="mRNA"
                    /db_xref="taxon:3760"
                    /clone="pfPpCesA1"
                    /tissue_type="fruit"
                    /clone_lib="Prunus persica fruit"
SE COUNT      239 a    215 c    234 g    272 t
IGIN
```

```
ignment Scores:
ed. No.:              2.02e-98        Length:         960
ore:                  1231.00         Matches:        225
rcent Similarity:     79.34%          Conservative:   40
st Local Similarity:  67.37%          Mismatches:     55
ery Match:            23.49%          Indels:         14
:                     9               Gaps:           3

-09-980-043-2 (1-978) x AB049602 (1-960)
```

```
Qy        504 AspArgTyrAlaAsnArgAsnValValPhePheAspValAsnMetLysGlyLeuAspGly 523
Db          1 GATAGATATGCTAATCGAAATGTTGTGTTCTTCGATATCAACATGAGAGGACTAGATGGG 60

Qy        524 IleGlnGlyProValTyrValGlyThrGlyCysValPheAsnArgGlnAlaLeuTyrGly 543
Db         61 ATCCAAGGACCAGTGTATGTTGGTACTGGATGTGTCTTCAATAGGCAAGCTTTGTATGGC 120

Qy        544 TyrGlyProProSerMetProSerLeuArgLysArgLysAspSerSerSerCysPheSer 563
Db        121 TATAATCCTCCAAAGGGTCCT----AAGCGCCCAAAGATGGTAAGCTGTGACTGC 171

Qy        564 CysCysCysProSerLysLysLysProAlaGlnAspProAlaGluValTyrArgAspAla 583
Db        172 TGCCCATGTTTTGGACGCGCAAGAAGCTTCCA----------AAGTAT 210

Qy        584 LysArgGluAspLeuAsnAlaAlaIlePheAsnLeuThrGluIleAspAsnTyrAspGlu 603
Db        211 TCCAAGCATGCTCAAATGGACAAGGTGCAAACCTACAAGGAGTGGATGAT----  161

Qy        604 HisGluArgSerMet...PheGluLysThrPheGlyLeuSerSer 623
Db        262 ---GACAAAGAGCTATTGATGT...AAATGAATTCGAGAAGAAATTCGGACAGTCCGCA 318

Qy        624 ValPheIleGluSerThrLeuMetGluAsnGlyGlyValProGluSerAlaAsnSerPro 643
Db        319 GTTTTTGTGACTTCAACTTTAATGGAACAAGGTGGTGTCCCTCCTTCCTCAAGCCCTGCG 378

Qy        644 ProPheIleLysGluAlaIleGlnValIleGlyCysGlyTyrGluGluLysThrGluTrp 663
Db        379 GCCATGCTTAAAGAAGCCATTCACGTAATCAGTTGCGGTTATGAAGACAAAACTGAATGG 438

Qy        664 GlyLysGlnIleGlyTrpIleTyrGlySerValThrGluAspIleLeuSerGlyPheLys 683
Db        439 GGTTTAGAGCTGGGTTGGATTTACGGGTCTATCACAGAGGATATCCTAACAGGTTTCAAG 498

Qy        684 MetHisCysArgGlyTrpArgSerIleTyrCysMetProValArgProAlaPheLysGly 703
Db        499 ATGCATTGCCGGGGTTGGAGGTCAATTTACTGTATGCCAGAGAGACCTGCATTCAAGGGA 558

Qy        704 SerAlaProIleAsnLeuSerAspArgLeuHisGlnValLeuArgTrpAlaLeuGlySer 723
Db        559 ACAGCTCCCATTAACTTGTCAGATCGGCTCAACCAGGTGCTTCGTTGGGCACTTGGTTCC 618

Qy        724 ValGluIlePhePheSerArgHisCysProLeuTrpTyrGlyPheGlyGlyGlyArgLeu 743
Db        619 ATTGAGATCTTTTTCAGTCGCCACAGTCCACTCTGGTATGGCTACAAAGGAGGGAAGCTC 678

Qy        744 LysTrpLeuGlnArgLeuAlaTyrIleAsnThrIleValTyrProPheThrSerLeuPro 763
Db        679 AAGTGGCTTGAGCGGTTCGCTTATGTCAACACTACTGTCTACCCTTTCACCTCCCTTCCT 738

Qy        764 LeuIleAlaTyrCysThrIleProAlaValCysLeuLeuThrGlyLysPheIleIlePro 783
Db        739 CTCCTTGCCTATTGTATCCTACCTGCCGTCTGCTTGCTCACCGATAAATTCATCATGCCA 798

Qy        784 ThrLeuSerAsnLeuAlaSerMetLeuPheLeuGlyLeuPheIleSerIleIleValThr 803
Db        799 TCGATAAGCACTTTTGCAAGTCTCTTCTTCATAGCTCTGTTCCTCTCAATCTTTGCCACC 858

Qy        804 AlaValLeuGluLeuArgTrpSerGlyValSerIleGluAspLeuTrpArgAsnGluGln 823
Db        859 GGCATTCTTGAGCTGCGATGGAGTGGAGTAAGCATCGAAGAATGGTGGAGAAATGAGCAA 918

Qy        824 PheTrpValIleGlyGlyValSerAlaHisLeuPheAlaVal 837
Db        919 TTCTGGGTGATTGGTGGTGTGTCAGCTCACCTCTTTGCTGTT 960
```

```
RESULT 12
BZ827769
LOCUS       BZ827769               1022 bp    DNA     linear   GSS 18-MAR-2003
DEFINITION  PUFHO17TB ZM_0.6_1.0_KB Zea mays genomic clone ZMMBTa345C09,
```

```
            genomic survey sequence.
ACCESSION   BZ827769
VERSION     BZ827769.1  GI:29048637
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 1022)
  AUTHORS   Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick
            ,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and Bennetzen,J.
  TITLE     Maize Genomics Consortium
  JOURNAL   Unpublished
COMMENT     Other_GSSs: PUFHO17TD
            Contact: Cathy Whitelaw
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-5843
            Fax: 301-838-0208
            Email: whitelaw@tigr.org
            Seq primer: TR
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1..1022
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBTa345C09"
                     /clone_lib="ZM_0.6_1.0_KB"
                     /note="Vector: pCR4-TOPO; Site_1: EcoRI; 0.6-1.0 kb high
                     CoT selected genomic DNA library"
BASE COUNT      210 a    247 c    254 g    311 t
ORIGIN

Alignment Scores:
Pred. No.:              2.51e-97        Length:         1022
Score:                  1219.00         Matches:        222
Percent Similarity:     80.12%          Conservative:   44
Best Local Similarity:  66.87%          Mismatches:     31
Query Match:            23.26%          Indels:         36
DB:                     29              Gaps:           5

US-09-980-043-2 (1-978) x BZ827769 (1-1022)

Qy        668 GlyTrpIleTyrGlySerValThrGluAspIleLeuSerGlyPheLysMetHisCysArg 687
              ||||||:|||||||||||||||||:|||   |||   |||||||||||||||||||:::
Db          5 GGTTGGATTTATGGATCAGTCACAGAAGATATTCTTACTGGTTTAAGATGCACTGCCAT 64

Qy        688 GlyTrpArgSerIleTyrCysMetProValArgProAlaPheLysGlySerAlaProIle 707
              ||||||:|||||||||||||||||:|||   |||   |||||||||||||||||||:::
Db         65 GGTTGGCGGTCAATTTACTGCATACCTAAACGGGCCGCCTTCAAAGGTTCCGCACCTCTC 124

Qy        708 AsnLeuSerAspArgLeuHisGlnValLeuArgTrpAlaLysGlySerValGluIlePhe 727
              ||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||||
Db        125 AATCTTTCCGATCGTCTTCACCAGGTTCTTCGGTGGGCTCTTGGTTCAATTGAAATTTTC 184

Qy        728 PheSerArgHisCysProLeuTrpTyrGlyPheGlyGlyArgLeuLysTrpLeuGln 747
              ||||||  ||||||||||||||||||:|||:|||||||  ||||:|:|||:::
Db        185 TTCAGCAACCACTGCCCTCTCTGGTATGGGTATGGTGGTGGA  -CTAAAGTTCCTGGAA 241

Qy        748 ArgLeuAlaTyrIleAsnThrIleValTyrProPheThrSerLeuProLeuIleAlaTyr 767
              |||  :::|||||||||:::|||||||:|||:::|||||:|:::|||||:::||||||
Db        242 AGGTTTTCGTACATTAACTCCATCGTATACCCTTGGACATCTATCCGCTCTTGGCCTAT 301

Qy        768 CysThrIleProAlaValCysLeuLeuThrGlyLysPheIleIleProThr--------- 784
              |||||:::||||||:::|||||||||||||||||||||||||||||||  |||
Db        302 TGCACATTGCCTGCCATCTGCTTGCTGACAGGGAAATTTATCACGCCAGA-GGTACCCAT 360

Qy        784 ------------------------------ ------  -------- -- --------------- 784

Db        361 TTATTCAGTCTCTAATTCCGTTGTATTGTGGTGTTGAACCGGTGGAAGATTATGACACCT 420

Qy        785 ----------------------------LeuSerAsnLeuAlaSerMetLeuPheLeu 794
                                           |||:::|||:::||||||:::   |||:::
Db        421 TACGATTACTATCATTCTCTTATCGTGCAGCTTAACAATGTTGCCAGCCTGTGGTTCATG 480

Qy        795 GlyLeuPheIleSerIleIleValThrAlaValLeuGluLeuArgTrpSerGlyValSer 814
              |||||||||   |||       |||::::::|||||||:::|||||||||||||||||
Db        481 TCACTTTTCATCTGCATTTTTGCTACGAGCATCCTGGAAATGAGATGGAGTGGTGTAGGC 540

Qy        815 IleGluAspLeuTrpArgAsnGluGlnPheTrpValIleGlyGlyValSerAlaHisLeu 834
              |||:::|||   ||||||||||||||||||||||||||||||||||||:::||||||
Db        541 ATCGATGACTGGTGGAGAAACGAGCAGTTTTGGGTCATTGGAGGCGTGTCTTCACATCTC 600

Qy        835 PheAlaValPheGlnGlyPheLeuLysMetLeuAlaGlyIleAspThrAsnPheThrVal 854
              ||||||||||||||||||   ||||||::::::||||||:::||||||:::|||||||||
Db        601 TTTGCTGTGTTCCAGGGACTCCTCAAGGTCATAGCTGGTGTAGACACGAGCTTCACTGTG 660

Qy        855 ThrAlaLysAlaAlaGluAspAlaGluPheGlyGluLeuTyrMetValLysTrpThrThr 874
              |||:::|||      :::|||   ||||||   |||||||||      ||||||||||
Db        661 ACATCCAAGGGCGGAGACGACGAGGAGTTCTCAGAGCTGTACACATTCAAATGGACGACC 720

Qy        875 LeuLeuIleProProThrThrLeuLeuIleIleAsnMetSerGly---CysAlaGlyPhe 893
              |||||||||||||||||||||||||:::::|||      |||      ||||||
Db        721 CTTCTGATACCTCCGACAACCCTGCTCCTACTGAACTTCATTGGAGTGGTAGCTGGCATC 780

Qy        894 SerAspAlaLeuAsnLysGlyTyrGluAlaTrpGlyProLeuPheGlyLysValPhePhe 913
              |||:::|||:::|||  |||||||||:::|||||||||||||||||||:::||||||
Db        781 TCCAATGCGATCAACAACGGATATGAATCATGGGGCCCCCTGTTCGGGAAGCTCTTCTTT 840

Qy        914 AlaPheTrpValIleLeuHisLeuTyrProPheLeuLysGlyLeuMetGlyArgGlnAsn 933
              |||||||||||||||:::|||||||||||||||||||||||||||:::||||||||||||
Db        841 GCATTTTGGGTGATCGTCCATCTTTACCCGTTCCTCAAGGGTCTGGTTGGGAGGCAGAAC 900

Qy        934 LeuThrProThrIleValValLeuTrpSerValLeuLeuAlaSerValPheSerLeuVal 953
              ||||||||||||||:::::|||||:::|||||||||||||:::|||||||||:::
Db        901 AGGACGCCAACGATTGTCATTGTCTGGTCCATCCTCCTGGCTTCGATCTTCTCGCTGCTT 960

Qy        954 TrpValLysIleAsnProPheValAsnLysValAsp 965
              ||||||:::     |||||:::   |||   |||
Db        961 TGGGTCCGGGATCGACCGTTCCTTGCGAAGGATGAT 996

RESULT 13
CB671068
LOCUS       CB671068                 792 bp    mRNA    linear   EST 09-APR-2003
DEFINITION  OSJNEe04H10.f OSJNEe Oryza sativa (japonica cultivar-group) cDNA
            clone OSJNEe04H10 5', mRNA sequence.
ACCESSION   CB671068
VERSION     CB671068.1  GI:29674793
KEYWORDS    EST.
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 792)
  AUTHORS   Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
            Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.
  TITLE     Large-scale identification of ESTs involved in the interaction
            between rice and Magnaporthe grisea
  JOURNAL   Unpublished
COMMENT     Contact: Rod Wing
            Arizona Genomics Institute
            University of Arizona
            Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
            85721-0088, USA
            Tel: 520 626 3967
            Fax: 520 621 9288
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: gta aaa cga cgg cca gtg
            BACKWARD: gga aac agc tat gac cat g
            Plate: 04  row: H  column: 10
            Seq primer: gta aaa cga cgg cca gtg.
```

```
ATURES             Location/Qualifiers
   source          1. .792
                   /organism="Oryza sativa (japonica cultivar-group)"
                   /mol_type="mRNA"
                   /cultivar="Nipponbare"
                   /db_xref="taxon:39947"
                   /clone="OSJNEe04H10"
                   /tissue_type="Leaf"
                   /dev_stage="3 week"
                   /lab_host="DH10B"
                   /clone_lib="OSJNEe"
                   /note="Vector: pBluescript II KS +; Site_1: EcoRI; Site_2:
                   XhoI; 24 hrs after innoculation with Rice Blast (70-15)"
SE COUNT      233 a    145 c    200 g    214 t
IGIN

ignment Scores:
ed. No.:              1.48e-95       Length:        792
ore:                  1197.50        Matches:       224
rcent Similarity:     92.42%         Conservative:  20
st Local Similarity:  84.85%         Mismatches:    20
ery Match:            22.85%         Indels:        3
:                     14             Gaps:          0

-09-980-043-2 (1-978) x CB671068 (1-792)

        281 AspTyrProValAspLysValSerCysTyrValSerAspAspGlyAlaAlaMetLeuSer 300
            ||||||||||||||||  ||||||||||||||||||||||||||||||||||||||||:::
          2 GATTATCCAGTTGATAATGTTTCATGCTATGTCTCTGATGATGGTGCTGCAATGCTGACA 61

        301 PheGluSerLeuValGluThrAlaGluPheAlaArgLysTrpValProPheCysLysLys 320
            ||||||:::|||  ||||||:::|||||||||:::||||||||||||||||||||||||
         62 TTTGAAGCATTATCCGAAACATCTGAATTTGCAAAGAAATGGGTTCCTTTCTGCAAAAAG 121

        321 PheSerIleGluProArgAlaProGluPheTyrPheSerGlnLysIleAspTyrLeuLys 340
            :::|||||||||||||||||||||||:::||||||  ||||||||||||||||||||||
        122 TACAGTATTGAACCACGTGCTCCAGAGTGGTACTTCCAACAGAAGATAGACTACCTGAAG 181

        341 AspLysValGlnProSerPheValLysGluArgArgAlaMetLysArgAspTyrGluGlu 360
            |||||||||  |||  ||||||:::|||||||||||||||||||||:::||||||||||
        182 GATAAGGTGGCACCATACTTTGTTAGGGAGAGGAGAGCAATGAAGAGAGAGTATGAGGAA 241

        361 TyrLysValArgValAsnAlaLeuValAlaLysAlaGlnLysThrProGluGluGlyTrp 380
            :::||||||||:::|||||||||||||||||||||||||||  ||||||||||||||||
        242 TTCAAGGTGAGAATCAATGCATTGGTTGCTAAAGCTCAGAAAGTTCCCGAGGAAGGTTGG 301

        381 ThrMetGlnAspGlyThrProTrpProGlyAsnAsnThrArgAspHisProGlyHisAsp 400
            |||||||||||||||||||||||||||||||||||  ||||||||||||||  |||||||
        302 ACAATGCAAGATGGAACTCCCTGGCCTGGAAACAATGTTCGCGATCACCCTGGAAT-GAT 360

        401 SerGlyLeuProTrpGluIleLeuGlyAlaArgAspIleGluGlyAsnGluLeuProArg 420
            ||||||||||||||  |||  ||||||||||||||||||||||||||||||||||||||
        361 TCAGGTCTTCCTTGGCCAA--AGTGGCGGCCATGACATCGAAGGTAATGAGCTACCTCGA 418

        421 LeuValTyrValSerArgGluLysArgProGlyTyrGlnHisHisLysLysAlaGlyAla 440
            |||||||||||||||||||||||||||||||||||  ||||||||||||||||||||||
        419 TTGGTGTATGTTTCAAGAGAAAAAAGGCCAGGCTATAACCATCACAAGAAGGCTGGTGCC 478

        441 GluAsnAlaLeuValArgValSerAlaValLeuThrAsnAlaProTyrIleLeuAsnVal 460
            |||||||||||||||||||||||||||||||||||||||||||||||:::||||||:::
        479 ATGAACGCATTGGTCCGAGTATCTGCTGTACTAACTAATGCTCCATACATGCTCAACTTG 538

        461 AspCysAspHisTyrValAsnAsnSerLysAlaValArgGluAlaMetCysIleLeuMet 480
            |||||||||||||||:::|||||||||||||||:::::||||||||||  :::|||||
        539 GATTGTGATCATTACATCAACAATAGCAAGGCTATAAAGGAAGCAATGTGTTTTATGATG 598

        481 AspProGlnValGlyArgAspValCysTyrValGlnPheProGlnArgPheAspGlyIle 500
            ||||||  ||||||:::  ||||||||||||||||||||||||||||||||||||||||
        599 GACCCTTTGGTTGGAAAGAAGGTTTGTTATGTGCAATTCCCTCAAAGGTTTGATGGGATT 658

        501 AspLysSerAspArgTyrAlaAsnArgAsnValValPhePheAspValAsnMetLysGly 520
            |||:::  ||||||||||||||||||||||||||||||||||:::||||||||||||||
        659 GATCGTCATGATCGATACGCCAACAGGAACGTGGTCTTCTTCGATATCAACATGAAAGGC 718

Qy      521 LeuAspGlyIleGlnGlyProValTyrValGlyThrGlyCysValPheAsnArgGlnAla 540
            ||||||||||||||||| |||:::|| ||||:|||||||||||||||||  ||||||||
Db      719 TTGGATGGTATCCAAGGCCCCATCTATGTTGGTACTGGATGTGTCTTTAC-AGGCAGGCA 777

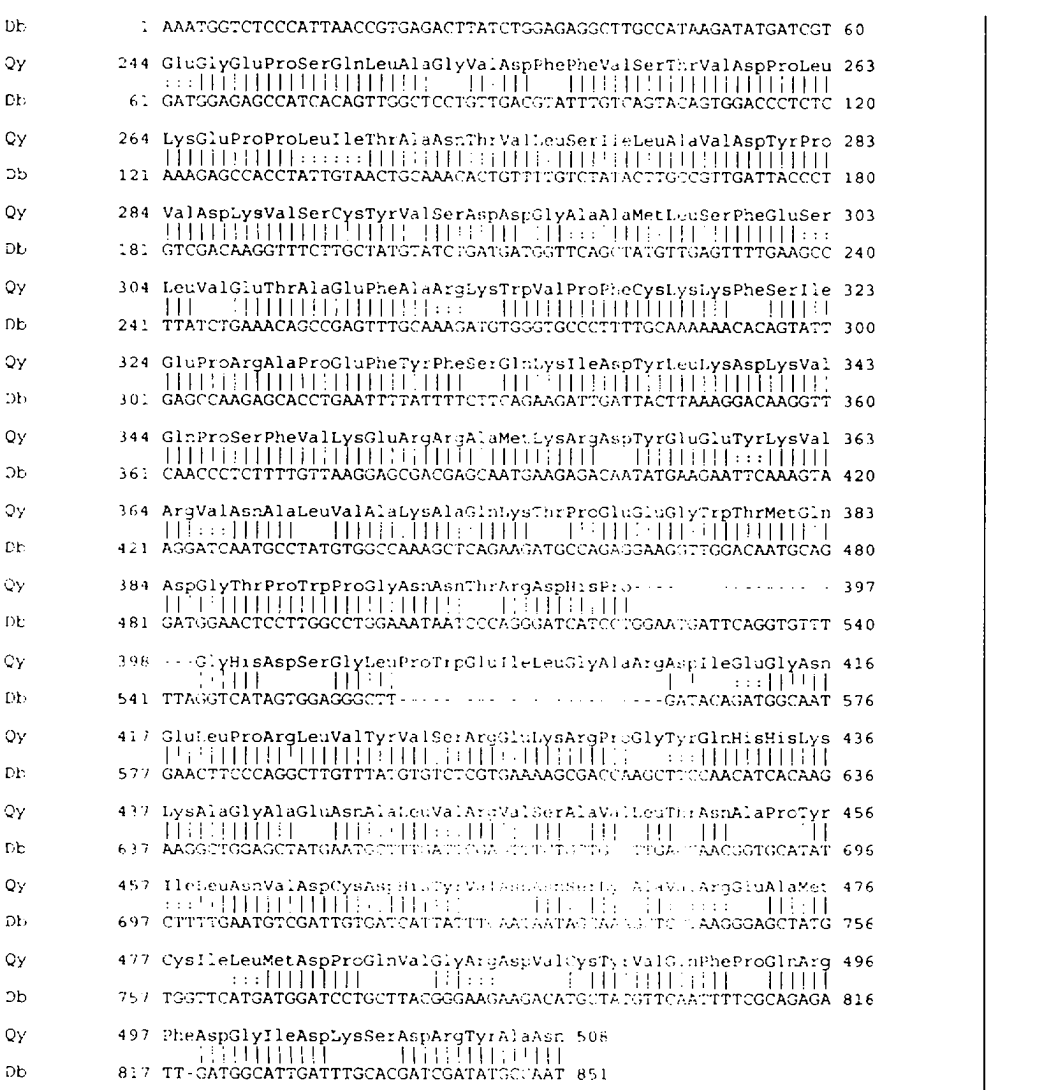
Qy      541 LeuTyrGlyTyr 552
            ||||||||||||
Db      778 TTGTATGGTTAT 789

RESULT 14
CB892280
LOCUS       CB892280                851 bp    mRNA    linear   EST 24-APR-2003
DEFINITION  EST649249 KV3 Medicago truncatula cDNA clone KV3-53B20, mRNA
            sequence.
ACCESSION   CB892280
VERSION     CB892280.1  GI:34099448
KEYWORDS    EST.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 851)
  AUTHORS   VandenBosch,K., Endre,G., Burr,J., Bertrand,P., Town,C.D., Van Aken
            ,S., Utterback,T., Cho,J. and Fraser,C.M.
  TITLE     ESTs from a cDNA of Medicago truncatula 72 h after Rhizobium
            inoculation, 2003
  JOURNAL   Unpublished
COMMENT     Contact: VandenBosch K
            Department of Plant Biology
            University of Minnesota
            220 BioSci Center, 1445 Gortner Ave. St  Paul, MN 55108, USA
            Tel: 612 624 2757
            Fax: 612 625 1738
            Email: kvandenb@tbs.umn.edu
            TIGR sequence name: MTEDP10TK
            More information is available at: www.medicago.org
            Seq primer: SKmod (CTA gAA CTA gtg gAT CC).
FEATURES             Location/Qualifiers
     source          1. .851
                     /organism="Medicago truncatula"
                     /mol_type="mRNA"
                     /cultivar="genotype A17"
                     /db_xref="taxon:3880"
                     /clone="KV3-53B20"
                     /tissue_type="Seedling roots"
                     /dev_stage="3 days post-inoculation with Sinorhizobium
                     meliloti"
                     /lab_host="E. coli strain XLOLR"
                     /clone_lib="KV3"
                     /note="Vector: pBluescript SK -; Site_1: EcoRI; Site_2:
                     XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA
                     was directionally ligated into the Unizap XR vector from
                     Stratagene and packaged using Gigapack III Gold packaging
                     extracts.  Plasmids containing cDNA inserts were excised
                     from the recombinant lambda-Zap phage using Ex-assist
                     helper phage and propagated in XLOLR cells."
BASE COUNT      238 a    159 c    204 g    250 t
ORIGIN

Alignment Scores:
Pred. No.:              1.01e-94       Length:        851
Score:                  1198.50        Matches:       226
Percent Similarity:     84.69%         Conservative:  21
Best Local Similarity:  77.40%         Mismatches:    30
Query Match:            22.68%         Indels:        16
DB:                     14             Gaps:          2

US-09-980-043-2 (1-978) x CB892280 (1-851)

Qy      224 LysTrpAsnProValAsnArgGluThrTyrIleGluArgLeuSerAlaArgTyrGluArg 243
            |||||||:::|||:::|||||||||||||||:::||||||||||:::  |||:|||:::|
```

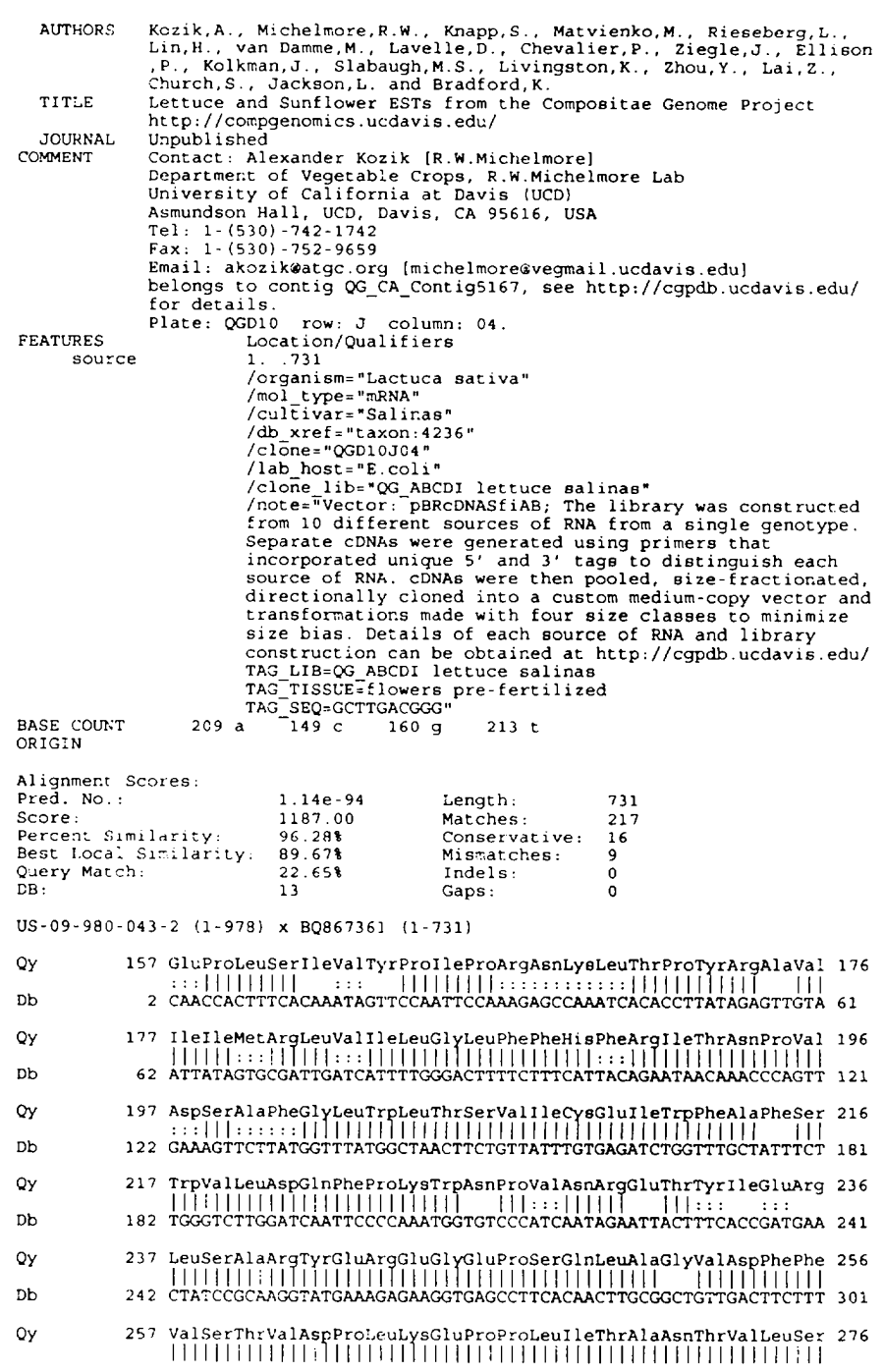


RESULT 15
BQ867361
LOCUS BQ867361 731 bp mRNA linear EST 14-AUG-2002
DEFINITION QGD10J04.yg.ab1 QG_ABCDI lettuce salinas Lactuca sativa cDNA clone QGD10J04, mRNA sequence.
ACCESSION BQ867361
VERSION BQ867361.1 GI:22253118
KEYWORDS EST.
SOURCE Lactuca sativa
ORGANISM Lactuca sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids; campanulids; Asterales; Asteraceae; Cichorioideae; Cichorieae; Lactuca.
REFERENCE 1 (bases 1 to 731)
AUTHORS Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L., Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegle,J., Ellison,P., Kolkman,J., Slabaugh,M.S., Livingston,K., Zhou,Y., Lai,Z., Church,S., Jackson,L. and Bradford,K.
TITLE Lettuce and Sunflower ESTs from the Compositae Genome Project http://compgenomics.ucdavis.edu/
JOURNAL Unpublished
COMMENT Contact: Alexander Kozik [R.W.Michelmore]
Department of Vegetable Crops, R.W.Michelmore Lab
University of California at Davis (UCD)
Asmundson Hall, UCD, Davis, CA 95616, USA
Tel: 1-(530)-742-1742
Fax: 1-(530)-752-9659
Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu]
belongs to contig QG_CA_Contig5167, see http://cgpdb.ucdavis.edu/ for details.
Plate: QGD10 row: J column: 04.
FEATURES Location/Qualifiers
source 1..731
/organism="Lactuca sativa"
/mol_type="mRNA"
/cultivar="Salinas"
/db_xref="taxon:4236"
/clone="QGD10J04"
/lab_host="E.coli"
/clone_lib="QG_ABCDI lettuce salinas"
/note="Vector: pBRcDNASfiAB; The library was constructed from 10 different sources of RNA from a single genotype. Separate cDNAs were generated using primers that incorporated unique 5' and 3' tags to distinguish each source of RNA. cDNAs were then pooled, size-fractionated, directionally cloned into a custom medium-copy vector and transformations made with four size classes to minimize size bias. Details of each source of RNA and library construction can be obtained at http://cgpdb.ucdavis.edu/
TAG_LIB=QG_ABCDI lettuce salinas
TAG_TISSUE=flowers pre-fertilized
TAG_SEQ=GCTTGACGGG"
BASE COUNT 209 a 149 c 160 g 213 t
ORIGIN

Alignment Scores:

Pred. No.:	1.14e-94	Length:	731
Score:	1187.00	Matches:	217
Percent Similarity:	96.28%	Conservative:	16
Best Local Similarity:	89.67%	Mismatches:	9
Query Match:	22.65%	Indels:	0
DB:	13	Gaps:	0

US-09-980-043-2 (1-978) x BQ867361 (1-731)



```
        302 GTGAGCACAGTTGACCCTTTGAAAGAGCCTCCTCTAATCACAGCCAATACTGTACTTTCA 361

        277 IleLeuAlaValAspTyrProValAspLysValSerCysTyrValSerAspAspGlyAla 296
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
        362 ATCCTAGCTGTTGATTACCCCGTTGATAAAGTCTCATGTTATGTCTCGGATGATGGTGCT 421

        297 AlaMetLeuSerPheGluSerLeuValGluThrAlaGluPheAlaArgLysTrpValPro 316
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
        422 GCAATGCTTTCGTTTGAATCCCTTGTTGAAACCGCGGAGTTTGCAAGGAAGTGGGTCCCG 481

        317 PheCysLysLysPheSerIleGluProArgAlaProGluPheTyrPheSerGlnLysIle 336
            |||||||||:::|||||||||||||||||||||||||||||||||||||||||||||||
        482 TTTTGTAAACGGTTTTCAATTGAGCCACGTGCACCTGAATTTTACTTCTCCCAAAAGATC 541

        337 AspTyrLeuLysAspLysValGlnProSerPheValLysGluArgArgAlaMetLysArg 356
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
        542 GATTACTTGAAAGATAAAGTCCAACCTTCATTTGTGAAGGAACGTAGAGCAATGAAACGG 601

        357 AspTyrGluGluTyrLysValArgValAsnAlaLeuValAlaLysAlaGlnLysThrPro 376
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
        602 GATTATGAGGAGTATAAAGTGAGGGTAAATGCTTTAGTGGCTAAAGCTCAGAAAACCCCG 661

        377 GluGluGlyTrpThrMetGlnAspGlyThrProTrpProGlyAsnAsnThrArgAspHis 396
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
        662 GAAGAAGGATGGACCATGCAAGATGGAACGCCGTGGCCGGGAAATAACACACGTGACCAC 721

        397 ProGly 398
            ||||||
        722 CCGGGA 727

earch completed: October 31, 2003, 17:42:35
ob time : 5891.74 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 29, 2003, 18:34:34 ; Search time 52.0213 Seconds
(without alignments)
3217.546 Million cell updates/sec

Title: US-09-980-043-2
Perfect score 978
Sequence: 1 MMESGAPICHTCGEQVGHDA.....PFVNKVLNTLVAETCISIDC 978

Scoring table: CLIGO
Gapop 60.0 , Gapext 60.0

Searched: 642050 seqs, 171146064 residues

Word size : 0

Total number of hits satisfying chosen parameters: 642050

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	61	6.2	974	10	US-09-838-539-6	Sequence 6, Appli
2	41	4.2	1007	12	US-10-209-059-28	Sequence 28, Appl
3	40	4.1	547	15	US-10-229-193-14	Sequence 14, Appl
4	40	4.1	1074	12	US-10-209-059-46	Sequence 46, Appl
5	40	4.1	1074	12	US-10-160-719-14	Sequence 14, Appl
6	40	4.1	1074	12	US-10-160-719-22	Sequence 22, Appl
7	40	4.1	1074	12	US-10-160-719-42	Sequence 42, Appl
8	39	4.0	685	10	US-09-900-237-31	Sequence 31, Appl
9	39	4.0	685	10	US-09-838-539-7	Sequence 7, Appli
10	36	3.7	793	10	US-09-900-237-18	Sequence 18, Appl
11	35	3.6	1065	15	US-10-229-193-10	Sequence 10, Appl
12	32	3.3	1084	15	US-10-229-193-8	Sequence 8, Appli
13	30	3.1	320	10	US-09-900-237-6	Sequence 6, Appli
14	30	3.1	740	10	US-09-900-237-24	Sequence 24, Appl
15	30	3.1	881	10	US-09-838-539-8	Sequence 8, Appli
16	30	3.1	1043	12	US-10-160-719-50	Sequence 50, Appl
17	30	3.1	1076	12	US-10-209-059-14	Sequence 14, Appl
18	30	3.1	1076	12	US-10-160-719-58	Sequence 58, Appl
19	30	3.1	1077	12	US-10-209-059-10	Sequence 10, Appl
20	30	3.1	1077	12	US-10-160-719-6	Sequence 6, Appli
21	30	3.1	1077	12	US-10-160-719-30	Sequence 30, Appl
22	30	3.1	1078	12	US-10-209-059-26	Sequence 26, Appl
23	30	3.1	1079	12	US-10-267-459-6	Sequence 6, Appli
24	30	3.1	1079	12	US-10-209-059-22	Sequence 22, Appl
25	27	2.8	927	12	US-10-209-059-30	Sequence 30, Appl
26	27	2.8	1080	10	US-09-900-237-30	Sequence 30, Appl
27	26	2.7	1091	10	US-09-900-237-26	Sequence 26, Appl
28	24	2.5	821	12	US-10-267-459-2	Sequence 2, Appli
29	24	2.5	821	12	US-10-209-059-6	Sequence 6, Appli
30	24	2.5	1039	10	US-09-900-237-14	Sequence 14, Appl
31	23	2.4	1075	12	US-10-209-059-2	Sequence 2, Appli
32	23	2.4	1075	12	US-10-160-719-10	Sequence 10, Appl
33	23	2.4	1075	12	US-10-160-719-34	Sequence 34, Appl
34	23	2.4	1075	12	US-10-160-719-54	Sequence 54, Appl
35	23	2.4	1081	15	US-10-229-193-6	Sequence 6, Appli
36	23	2.4	1081	15	US-10-229-193-12	Sequence 12, Appl
37	22	2.2	701	10	US-09-900-237-32	Sequence 32, Appl
38	22	2.2	1065	10	US-09-900-237-33	Sequence 33, Appl
39	22	2.2	1094	12	US-10-209-059-18	Sequence 18, Appl
40	22	2.2	1094	12	US-10-160-719-26	Sequence 26, Appl
41	22	2.2	1094	12	US-10-160-719-46	Sequence 46, Appl
42	22	2.2	1165	10	US-09-900-237-8	Sequence 8, Appli
43	21	2.1	629	15	US-10-229-193-2	Sequence 2, Appli
44	20	2.0	431	10	US-09-900-237-28	Sequence 28, Appl
45	20	2.0	1059	12	US-10-209-059-42	Sequence 42, Appl

ALIGNMENTS

RESULT 1
US-09-838-539-6
; Sequence 6, Application US/09838539
; Patent No. US20020129401A1
; GENERAL INFORMATION:
; APPLICANT: Stalker, D. et al.
; TITLE OF INVENTION: Plant Cellulose Synthase and Promoter
; TITLE OF INVENTION: Sequences
; FILE REFERENCE: 15621/03/US
; CURRENT APPLICATION NUMBER: US/09/838,539
; CURRENT FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/029,987
; PRIOR FILING DATE: 1996-10-29
; PRIOR APPLICATION NUMBER: 08/960,048
; PRIOR FILING DATE: 1997-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 974
; TYPE: PRT
; ORGANISM: Gossypim hirsutum
US-09-838-539-6

Query Match 6.2%; Score 61; DB 10; Length 974;
Best Local Similarity 100.0%; Pred. No. 9.7e-52;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy        799 SIIVTAVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQGFLKMLAGIDTNFTVTAKA 858
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        794 SIIVTAVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQGFLKMLAGIDTNFTVTAKA 853

Qy        859 A 859
              |
Db        854 A 854
```

RESULT 2
US-10-209-059-28

Sequence 28, Application US/10209059
Publication No. US20030163838A1
GENERAL INFORMATION:
APPLICANT: Dhugga, Kanwarpal S.
APPLICANT: Wang, Haiyin
TITLE OF INVENTION: Maize Cellulose Synthases and Uses
TITLE OF INVENTION: Thereof
FILE REFERENCE: 0864R2
CURRENT APPLICATION NUMBER: US/10/209,059
CURRENT FILING DATE: 2002-07-31
PRIOR APPLICATION NUMBER: 60/096,822
PRIOR FILING DATE: 1998-08-17
PRIOR APPLICATION NUMBER: 09/371,383
PRIOR FILING DATE: 1999-08-06
PRIOR APPLICATION NUMBER: 09/550,483
PRIOR FILING DATE: 2000-04-14
NUMBER OF SEQ ID NOS: 52
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 28
LENGTH: 1007
TYPE: PRT
ORGANISM: Zea mays
-10-209-059-28

Query Match 4.2%; Score 41; DB 12; Length 1007;
Best Local Similarity 100.0%; Pred. No. 9.9e-32;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

681 GFKMHCRGWRSIYCMPVRPAFKGSAPINLSDRLHQVLRWAL 721
686 GFKMHCRGWRSIYCMPVRPAFKGSAPINLSDRLHQVLRWAL 726

SULT 3
-10-229-193-14
Sequence 14, Application US/10229193
Publication No. US20030126643A1
GENERAL INFORMATION:
APPLICANT: Arioli, Antonio Williamson, Richard E.
APPLICANT: Betzner, Andreas S.
APPLICANT: Peng, Liangcai
TITLE OF INVENTION: Manipulation of Cellulose and/or Beta-1.4-Glucan
FILE REFERENCE: 96-98A
CURRENT APPLICATION NUMBER: US/10/229,193
CURRENT FILING DATE: 2002-08-26
PRIOR APPLICATION NUMBER: 09/221,013
PRIOR FILING DATE: 1998-12-23
PRIOR APPLICATION NUMBER: PCT/AU97/00402
PRIOR FILING DATE: 1997-06-24
PRIOR APPLICATION NUMBER: AU PO0699
PRIOR FILING DATE: 1996-06-27
NUMBER OF SEQ ID NOS: 36
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 14
LENGTH: 547
TYPE: PRT
ORGANISM: Oryza sativa
-10-229-193-14

Query Match 4.1%; Score 40; DB 15; Length 547;
Best Local Similarity 100.0%; Pred. No. 5.9e-31;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 256 FVSTVDPLKEPPLITANTVLSILAVDYPVDKVSCYVSDDG 295
D 354 FVSTVDPLKEPPLITANTVLSILAVDYPVDKVSCYVSDDG 393

CSULT 4
S-10-209-059-46
Sequence 46, Application US/10209059
Publication No. US20030163838A1
GENERAL INFORMATION:
; APPLICANT: Dhugga, Kanwarpal S.
; APPLICANT: Wang, Haiyin
; TITLE OF INVENTION: Maize Cellulose Synthases and Uses
; TITLE OF INVENTION: Thereof
; FILE REFERENCE: 0864R2
; CURRENT APPLICATION NUMBER: US/10/209,059
; CURRENT FILING DATE: 2002-07-31
; PRIOR APPLICATION NUMBER: 60/096,822
; PRIOR FILING DATE: 1998-08-17
; PRIOR APPLICATION NUMBER: 09/371,383
; PRIOR FILING DATE: 1999-08-06
; PRIOR APPLICATION NUMBER: 09/550,483
; PRIOR FILING DATE: 2000-04-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 46
; LENGTH: 1074
; TYPE: PRT
; ORGANISM: Zea mays
US-10-209-059-46

Query Match 4.1%; Score 40; DB 12; Length 1074;
Best Local Similarity 100.0%; Pred. No. 1.1e-30;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 256 FVSTVDPLKEPPLITANTVLSILAVDYPVDKVSCYVSDDG 295
Db 352 FVSTVDPLKEPPLITANTVLSILAVDYPVDKVSCYVSDDG 391

RESULT 5
US-10-160-719-14
; Sequence 14, Application US/10160719
; Publication No. US20030167528A1
; GENERAL INFORMATION:
; APPLICANT: Kanwarpal S. Dhugga
; APPLICANT: Timothy G. Helentjaris
; APPLICANT: Benjamin A. Bowen
; APPLICANT: Xun Wang
; TITLE OF INVENTION: Maize Cellulose Synthases and Uses
; TITLE OF INVENTION: Thereof
; FILE REFERENCE: 0864
; CURRENT APPLICATION NUMBER: US/10/160,719
; CURRENT FILING DATE: 2002-06-03
; PRIOR APPLICATION NUMBER: US/09/371,383A
; PRIOR FILING DATE: 1999-08-06
; PRIOR APPLICATION NUMBER: US 60/096,822
; PRIOR FILING DATE: 1998-08-17
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 14
; LENGTH: 1074
; TYPE: PRT
; ORGANISM: Zea mays
US-10-160-719-14

Query Match 4.1%; Score 40; DB 12; Length 1074;
Best Local Similarity 100.0%; Pred. No. 1.1e-30;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 256 FVSTVDPLKEPPLITANTVLSILAVDYPVDKVSCYVSDDG 295
Db 352 FVSTVDPLKEPPLITANTVLSILAVDYPVDKVSCYVSDDG 391

RESULT 6
US-10-160-719-22
; Sequence 22, Application US/10160719
; Publication No. US20030167528A1
; GENERAL INFORMATION:
; APPLICANT: Kanwarpal S. Dhugga
; APPLICANT: Timothy G. Helentjaris
; APPLICANT: Benjamin A. Bowen

```
;  APPLICANT:  Xun Wang
;  TITLE OF INVENTION: Maize Cellulose Synthases and Uses
;  TITLE OF INVENTION:  Thereof
;  FILE REFERENCE: 0864
;  CURRENT APPLICATION NUMBER: US/10/160,719
;  CURRENT FILING DATE:  2002-06-03
;  PRIOR APPLICATION NUMBER: US/09/371,383A
;  PRIOR FILING DATE: 1999-08-06
;  PRIOR APPLICATION NUMBER: US 60/096,822
;  PRIOR FILING DATE: 1998-08-17
;  NUMBER OF SEQ ID NOS: 60
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 22
;   LENGTH: 1074
;   TYPE: PRT
;   ORGANISM: Zea mays
US-10-160-719-22

  Query Match             4.1%;  Score 40;  DB 12;  Length 1074;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-30;
  Matches   40;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        256 FVSTVDPLKEPPLITANTVLSILAVDYPVDKVSCYVSDDG 295
              ||||||||||||||||||||||||||||||||||||||||
Db        352 FVSTVDPLKEPPLITANTVLSILAVDYPVDKVSCYVSDDG 391


RESULT 7
US-10-160-719-42
; Sequence 42, Application US/10160719
; Publication No. US20030167528A1
; GENERAL INFORMATION:
;  APPLICANT: Kanwarpal S. Dhugga
;  APPLICANT:  Timothy G. Helentjaris
;  APPLICANT:  Benjamin A.  Bowen
;  APPLICANT:  Xun Wang
;  TITLE OF INVENTION: Maize Cellulose Synthases and Uses
;  TITLE OF INVENTION:  Thereof
;  FILE REFERENCE: 0864
;  CURRENT APPLICATION NUMBER: US/10/160,719
;  CURRENT FILING DATE:  2002-06-03
;  PRIOR APPLICATION NUMBER: US/09/371,383A
;  PRIOR FILING DATE: 1999-08-06
;  PRIOR APPLICATION NUMBER: US 60/096,822
;  PRIOR FILING DATE: 1998-08-17
;  NUMBER OF SEQ ID NOS: 60
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 42
;   LENGTH: 1074
;   TYPE: PRT
;   ORGANISM: Zea mays
US-10-160-719-42

  Query Match             4.1%;  Score 40;  DB 12;  Length 1074;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-30;
  Matches   40;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        256 FVSTVDPLKEPPLITANTVLSILAVDYPVDKVSCYVSDDG 295
              ||||||||||||||||||||||||||||||||||||||||
Db        352 FVSTVDPLKEPPLITANTVLSILAVDYPVDKVSCYVSDDG 391


RESULT 8
US-09-900-237-31
; Sequence 31, Application US/09900237
; Patent No. US20020120124A1
; GENERAL INFORMATION:
;  APPLICANT: Allen, Stephen
;  TITLE OF INVENTION: Plant Cellulose Synthases
;  FILE REFERENCE: BB1170 US CIP
;  CURRENT APPLICATION NUMBER: US/09/900,237
;  CURRENT FILING DATE:  2001-07-06
;  PRIOR APPLICATION NUMBER: 60/092,844
;  PRIOR FILING DATE: 1998-07-14
;  PRIOR APPLICATION NUMBER: PCT/US99/15871
;  PRIOR FILING DATE: 1999-07-13
;  PRIOR APPLICATION NUMBER: 09/720383
;  PRIOR FILING DATE: 2000-12-21
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 31
;   LENGTH: 685
;   TYPE: PRT
;   ORGANISM: Gossypium hirsutum
US-09-900-237-31

  Query Match             4.0%;  Score 39;  DB 10;  Length 685;
  Best Local Similarity   100.0%;  Pred. No. 7.1e-30;
  Matches   39;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        417 ELPRLVYVSREKRPGYQHHKKAGAENALVRVSAVLTNAP 455
              |||||||||||||||||||||||||||||||||||||||
Db        105 ELPRLVYVSREKRPGYQHHKKAGAENALVRVSAVLTNAP 143


RESULT 9
US-09-838-539-7
; Sequence 7, Application US/09838539
; Patent No. US20020129401A1
; GENERAL INFORMATION:
;  APPLICANT: Stalker, D. et al.
;  TITLE OF INVENTION: Plant Cellulose Synthase and Promoter
;  TITLE OF INVENTION:  Sequences
;  FILE REFERENCE: 15621/03/US
;  CURRENT APPLICATION NUMBER: US/09/838,539
;  CURRENT FILING DATE:  2001-04-18
;  PRIOR APPLICATION NUMBER: 60/029,987
;  PRIOR FILING DATE: 1996-10-29
;  PRIOR APPLICATION NUMBER: 08/960,048
;  PRIOR FILING DATE: 1997-10-29
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7
;   LENGTH: 685
;   TYPE: PRT
;   ORGANISM: Gossypium hirsutum
US-09-838-539-7

  Query Match             4.0%;  Score 39;  DB 10;  Length 685;
  Best Local Similarity   100.0%;  Pred. No. 7.1e-30;
  Matches   39;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        417 ELPRLVYVSREKRPGYQHHKKAGAENALVRVSAVLTNAP 455
              |||||||||||||||||||||||||||||||||||||||
Db        106 ELPRLVYVSREKRPGYQHHKKAGAENALVRVSAVLTNAP 144


RESULT 10
US-09-900-237-18
; Sequence 18, Application US/09900237
; Patent No. US20020120124A1
; GENERAL INFORMATION:
;  APPLICANT: Allen, Stephen
;  TITLE OF INVENTION: Plant Cellulose Synthases
;  FILE REFERENCE: BB1170 US CIP
;  CURRENT APPLICATION NUMBER: US/09/900,237
;  CURRENT FILING DATE:  2001-07-06
;  PRIOR APPLICATION NUMBER: 60/092,844
;  PRIOR FILING DATE: 1998-07-14
;  PRIOR APPLICATION NUMBER: PCT/US99/15871
;  PRIOR FILING DATE: 1999-07-13
;  PRIOR APPLICATION NUMBER: 09/720383
;  PRIOR FILING DATE: 2000-12-21
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 18
```

LENGTH: 793
TYPE: PRT
ORGANISM: Glycine max
S-09-900-237-18

Query Match 3.7%; Score 36; DB 10; Length 793;
Best Local Similarity 100.0%; Pred. No. 8.1e-27;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

y 698 RPAFKGSAPINLSDRLHQVLRWALGSVEIFFSRHCP 733
||||||||||||||||||||||||||||||||||
o 513 RPAFKGSAPINLSDRLHQVLRWALGSVEIFFSRHCP 548

ESULT 11
S-10-229-193-10
Sequence 10, Application US/10229193
Publication No. US20030126643A1
GENERAL INFORMATION:
APPLICANT: Arioli, Antonio Williamson, Richard E.
APPLICANT: Betzner, Andreas S.
APPLICANT: Peng, Liangcai
TITLE OF INVENTION: Manipulation of Cellulose and/or Beta 1.4-Glucan
FILE REFERENCE: 96-98A
CURRENT APPLICATION NUMBER: US/10/229,193
CURRENT FILING DATE: 2002-08-26
PRIOR APPLICATION NUMBER: 09/221,013
PRIOR FILING DATE: 1998-12-23
PRIOR APPLICATION NUMBER: PCT/AU97/00402
PRIOR FILING DATE: 1997-06-24
PRIOR APPLICATION NUMBER: AU PO0699
PRIOR FILING DATE: 1996-06-27
NUMBER OF SEQ ID NOS: 36
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 10
LENGTH: 1065
TYPE: PRT
ORGANISM: Arabidopsis thaliana
S-10-229-193-10

Query Match 3.6%; Score 35; DB 15; Length 1065;
Best Local Similarity 100.0%; Pred. No. 1e-25;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

y 270 TANTVLSILAVDYPVDKVSCYVSDDGAAMLSFESL 304
|||||||||||||||||||||||||||||||||||
o 355 TANTVLSILAVDYPVDKVSCYVSDDGAAMLSFESL 389

ESULT 12
S-10-229-193-8
Sequence 8, Application US/10229193
Publication No. US20030126643A1
GENERAL INFORMATION:
APPLICANT: Arioli, Antonio Williamson, Richard E.
APPLICANT: Betzner, Andreas S.
APPLICANT: Peng, Liangcai
TITLE OF INVENTION: Manipulation of Cellulose and/or Beta-1.4-Glucan
FILE REFERENCE: 96-98A
CURRENT APPLICATION NUMBER: US/10/229,193
CURRENT FILING DATE: 2002-08-26
PRIOR APPLICATION NUMBER: 09/221,013
PRIOR FILING DATE: 1998-12-23
PRIOR APPLICATION NUMBER: PCT/AU97/00402
PRIOR FILING DATE: 1997-06-24
PRIOR APPLICATION NUMBER: AU PO0699
PRIOR FILING DATE: 1996-06-27
NUMBER OF SEQ ID NOS: 36
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 8
LENGTH: 1084
TYPE: PRT
ORGANISM: Arabidopsis thaliana
US-10-229-193-8

Query Match [illegible]; Score 32; DB 15; Length 1084;
Best Local Similarity [illegible]; Pred. No. 1.3e-22;
Matches 32; Conservative [illegible]; Mismatches [illegible]; Indels 0; Gaps 0;

Qy 256 FVSTVDPLFEPPLIANTVLSILAVDYPVEKV 287
||:|||| ||.: ||. ||| .||: ||
Db 359 FVSTVDPLFEPPLIANTVLSILAVDYPVEKV 390

RESULT 13
US-09-900-237-6
; Sequence 6, Application US [illegible]
; Patent No. US20020120124A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Stephen
; TITLE OF INVENTION: Plant Cellulose Synthases
; FILE REFERENCE: BB1170 US CIP
; CURRENT APPLICATION NUMBER: [illegible]
; CURRENT FILING DATE: [illegible]
; PRIOR APPLICATION NUMBER: [illegible]
; PRIOR FILING DATE: 1998-[illegible]
; PRIOR APPLICATION NUMBER: [illegible]
; PRIOR FILING DATE: 1999-[illegible]
; PRIOR APPLICATION NUMBER: 09/720,383
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 6
; LENGTH: 320
; TYPE: PRT
; ORGANISM: Zea mays
US-09-900-237-6

Query Match 3.1%; Score 30; DB 10; Length 320;
Best Local Similarity 100.0%; Pred. No. 3.7e-21;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 698 RPAFKGSAPINLSDRLHQVLRWALGSVEIF 727
||||||:|||||'||| ||:..||
Db 40 RPAFKGSAPINLSDRLHQVLRWALGSVEIF 69

RESULT 14
US-09-900-237-24
; Sequence 24, Application US/09900237
; Patent No. US20020120124A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Stephen
; TITLE OF INVENTION: Plant Cellulose Synthases
; FILE REFERENCE: BB1170 US CIP
; CURRENT APPLICATION NUMBER: US/09/900,237
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: 60/092,844
; PRIOR FILING DATE: 1998-07-14
; PRIOR APPLICATION NUMBER: PCT/US99/15971
; PRIOR FILING DATE: 1999-07-13
; PRIOR APPLICATION NUMBER: 09/720383
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 24
; LENGTH: 740
; TYPE: PRT
; ORGANISM: Picramnia pentandra
US-09-900-237-24

Query Match 3.1%; Score 30; DB 10; Length 740;
Best Local Similarity 100.0%; Pred. No. 7.6e-21;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 698 RPAFKGSAPINLSDRLHQVLRWALGSVEIF 727

```
              ||||||||||||||||||||||||:||||:|
Db        462 RPAFKGSAPINLSDRLHQVLRWALGSVEIF 491


RESULT 15
US-09-838-539-8
; Sequence 8, Application US/09838539
; Patent No. US20020129401A1
; GENERAL INFORMATION:
;  APPLICANT: Stalker, D. et al.
;  TITLE OF INVENTION: Plant Cellulose Synthase and Promoter
;  TITLE OF INVENTION:  Sequences
;  FILE REFERENCE: 15621/03/US
;  CURRENT APPLICATION NUMBER: US/09/838,539
;  CURRENT FILING DATE:  2001-04-18
;  PRIOR APPLICATION NUMBER: 60/029,997
;  PRIOR FILING DATE: 1996-10-29
;  PRIOR APPLICATION NUMBER: 08/960,049
;  PRIOR FILING DATE: 1997-10-29
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8
;   LENGTH: 881
;   TYPE: PRT
;   ORGANISM: Oryzae sativa
US-09-838-539-8

  Query Match             3.1%;  Score 30;  DB 10;  Length 881;
  Best Local Similarity   100.0%;  Pred. No. 8.8e-21;
  Matches   30;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        270 TANTVLSILAVDYPVDKVSCYVSDDGAAML 299
              ||||||||||||||||||||||||||||||
Db        172 TANTVLSILAVDYPVDKVSCYVSDDGAAML 201


Search completed: October 29, 2003, 14:38:32
Job time : 53.0213 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 29, 2003, 18:33:14 ; Search time 28.3752 Seconds
(without alignments)
1458.314 Million cell updates/sec

Title: US-09-980-043-2
Perfect score: 978
Sequence: 1 MMESGAPICHTCGEQVGHDA.....PFVNKVINTLVAETCISIDC 978

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 328717 seqs, 42310858 residues

Word size : 0

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post processing: Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	61	6.2	974	3	US-08-960-048-6	Sequence 6, Appli
2	61	6.2	974	4	US-09-838-586-6	Sequence 6, Appli
3	40	4.1	547	4	US-09-221-013A-14	Sequence 14, Appl
4	39	4.0	685	3	US-08-960-048-7	Sequence 7, Appli
5	39	4.0	685	4	US-09-838-586-7	Sequence 7, Appli
6	35	3.6	1065	4	US-09-221-013A-10	Sequence 10, Appl
7	32	3.3	1084	4	US-09-221-013A-8	Sequence 8, Appli
8	30	3.1	881	3	US-08-960-048-8	Sequence 8, Appli
9	30	3.1	881	4	US-09-838-586-8	Sequence 8, Appli
10	23	2.4	1081	4	US-09-221-013A-6	Sequence 6, Appli
11	23	2.4	1081	4	US-09-221-013A-12	Sequence 12, Appl
12	21	2.1	629	4	US-09-221-013A-2	Sequence 2, Appli
13	8	0.8	728	4	US-09-134-001C-4266	Sequence 4266, Ap
14	7	0.7	76	4	US-09-107-532A-6959	Sequence 6959, Ap
15	7	0.7	124	4	US-09-134-001C-4274	Sequence 4274, Ap
16	7	0.7	129	4	US-09-252-991A-28493	Sequence 28493, A
17	7	0.7	199	4	US-09-252-991A-32787	Sequence 32787, A
18	7	0.7	207	4	US-09-328-352-5708	Sequence 5708, Ap
19	7	0.7	220	4	US-09-844-497-2	Sequence 2, Appli
20	7	0.7	236	2	US-08-738-413B-10	Sequence 10, Appl
21	7	0.7	238	4	US-09-252-991A-24451	Sequence 24451, A
22	7	0.7	281	4	US-09-107-532A-4985	Sequence 4985, Ap
23	7	0.7	312	2	US-08-518-862C-2	Sequence 2, Appli
24	7	0.7	354	1	US-07-759-568-2	Sequence 2, Appli
25	7	0.7	359	4	US-09-336-643A-2	Sequence 2, Appli
26	7	0.7	400	4	US-09-252-991A-27422	Sequence 27422, A
27	7	0.7	406	4	US-09-328-352-7690	Sequence 7690, Ap
28	7	0.7	408	4	US-09-328-352-6550	Sequence 6550, Ap
29	7	0.7	412	2	US-08-741-134-2	Sequence 2, Appli
30	7	0.7	458	4	US-09-134-001C-4422	Sequence 4422, Ap
31	7	0.7	522	4	US-09-446-301A-49	Sequence 49, Appl
32	7	0.7	522	4	US-09-328-352-8107	Sequence 8107, Ap
33	7	0.7	522	4	US-09-099-932-49	Sequence 49, Appl
34	7	0.7	527	4	US-09-252-991A-23709	Sequence 23709, A
35	7	0.7	582	3	US-09-413-814-83	Sequence 83, Appl
36	7	0.7	692	4	US-09-252-991A-25793	Sequence 25793, A
37	7	0.7	738	3	US-08-864-038A-3	Sequence 3, Appli
38	7	0.7	889	1	US-08-118-101A-4	Sequence 4, Appli
39	7	0.7	1070	4	US-09-107-532A-3716	Sequence 3716, Ap
40	7	0.7	1106	4	US-09-252-991A-22239	Sequence 22239, A
41	7	0.7	1184	1	US-08-446-038B-20	Sequence 20, Appl
42	7	0.7	1184	1	US-08-446-010B-20	Sequence 20, Appl
43	7	0.7	1184	2	US-08-805-445-20	Sequence 20, Appl
44	7	0.7	1184	2	US-08-064-067D-20	Sequence 20, Appl
45	7	0.7	1184	2	US-09-066-208-20	Sequence 20, Appl

ALIGNMENTS

RESULT 1
US-08-960-048-6
; Sequence 6, Application US/08960048C
; Patent No. 6271443
; GENERAL INFORMATION:
; APPLICANT: Stalker, D. et al.
; TITLE OF INVENTION: Plant Cellulose Synthase and Promoter
; TITLE OF INVENTION: Sequences
; FILE REFERENCE: 15621/01/US
; CURRENT APPLICATION NUMBER: US/08/960,048C
; CURRENT FILING DATE: 1997-10-29
; PRIOR APPLICATION NUMBER: 60/029,987
; PRIOR FILING DATE: 1996-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 974
; TYPE: PRT
; ORGANISM: Gossypim hirsutum
US-08-960-048-6

Query Match 6.2%; Score 61; DB 3; Length 974;
Best Local Similarity 100.0%; Pred. No. 1.4e-53;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy        799 SIIVTAVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQGFLKMLAGIDTNFTVTAKA 858
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        794 SIIVTAVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQGFLKMLAGIDTNFTVTAKA 853

Qy        859 A 859
              |
Db        854 A 854
```

RESULT 2
US-09-838-586-6
; Sequence 6, Application US/09838586
; Patent No. 6576818
; GENERAL INFORMATION:
; APPLICANT: Stalker, D. et al.
; TITLE OF INVENTION: Plant Cellulose Synthase and Promoter
; TITLE OF INVENTION: Sequences
; FILE REFERENCE: 15621/02/US
; CURRENT APPLICATION NUMBER: US/09/838,586
; CURRENT FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/029,987
; PRIOR FILING DATE: 1996-10-29
; PRIOR APPLICATION NUMBER: 08/960,048
; PRIOR FILING DATE: 1997-10-29
; NUMBER OF SEQ ID NOS: 12

SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 6
LENGTH: 974
TYPE: PRT
ORGANISM: Gossypim hirsutum
-09-838-586-6

Query Match 6.2%; Score 61; DB 4; Length 974;
Best Local Similarity 100.0%; Pred. No. 1.4e-53;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

799 SIIVTAVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQGFLKMLAGIDTNFTVTAKA 858
794 SIIVTAVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQGFLKMLAGIDTNFTVTAKA 853

859 A 859
854 A 854

SULT 3
-09-221-013A-14
Sequence 14, Application US/09221013A
Patent No. 6495740
GENERAL INFORMATION:
APPLICANT: Arioli, Antonio
APPLICANT: Williamson, Richard E.
APPLICANT: Betzner, Andreas S.
APPLICANT: Peng, Liangcai
TITLE OF INVENTION: Manipulation of cellulose and/or Beta-1,4-glucan
FILE REFERENCE: 96-98
CURRENT APPLICATION NUMBER: US/09/221,013A
CURRENT FILING DATE: 1998-12-23
PRIOR APPLICATION NUMBER: PCT/AU97/00402
PRIOR FILING DATE: 1997-06-24
PRIOR APPLICATION NUMBER: AU PO0699
PRIOR FILING DATE: 1996-06-27
NUMBER OF SEQ ID NOS: 37
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 14
LENGTH: 547
TYPE: PRT
ORGANISM: Oryza sativa
-09-221-013A-14

Query Match 4.1%; Score 40; DB 4; Length 547;
Best Local Similarity 100.0%; Pred. No. 3e-32;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

256 FVSTVDPLKEPPLITANTVLSILAVDYPVDKVSCYVSDDG 295
354 FVSTVDPLKEPPLITANTVLSILAVDYPVDKVSCYVSDDG 393

SULT 4
-08-960-048-7
Sequence 7, Application US/08960048C
Patent No. 6271443
GENERAL INFORMATION:
APPLICANT: Stalker, D. et al.
TITLE OF INVENTION: Plant Cellulose Synthase and Promoter
TITLE OF INVENTION: Sequences
FILE REFERENCE: 15621/01/US
CURRENT APPLICATION NUMBER: US/08/960,048C
CURRENT FILING DATE: 1997-10-29
PRIOR APPLICATION NUMBER: 60/029,987
PRIOR FILING DATE: 1996-10-29
NUMBER OF SEQ ID NOS: 12
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 7
LENGTH: 685
TYPE: PRT
ORGANISM: Gossypium hirsutum

US-08-960-048-7

Query Match 4.0%; Score 39; DB 3; Length 685;
Best Local Similarity 100.0%; Pred. No. 3.9e-31;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 417 ELPRLVYVSREKRPGYQHHKKAGAENALVRVSAVLTNAP 455
Db 106 ELPRLVYVSREKRPGYQHHKKAGAENALVRVSAVLTNAP 144

RESULT 5
US-09-838-586-7
; Sequence 7, Application US/09838586
; Patent No. 6576618
; GENERAL INFORMATION:
; APPLICANT: Stalker, D. et al.
; TITLE OF INVENTION: Plant Cellulose Synthase and Promoter
; TITLE OF INVENTION: Sequences
; FILE REFERENCE: 15621/02/US
; CURRENT APPLICATION NUMBER: US/09/838,586
; CURRENT FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/029,987
; PRIOR FILING DATE: 1996-10-29
; PRIOR APPLICATION NUMBER: 08/960,048
; PRIOR FILING DATE: 1997-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 685
; TYPE: PRT
; ORGANISM: Gossypium hirsutum
US-09-838-586-7

Query Match 4.0%; Score 39; DB 4; Length 685;
Best Local Similarity 100.0%; Pred. No. 3.9e-31;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 417 ELPRLVYVSREKRPGYQHHKKAGAENALVRVSAVLTNAP 455
Db 106 ELPRLVYVSREKRPGYQHHKKAGAENALVRVSAVLTNAP 144

RESULT 6
US-09-221-013A-10
; Sequence 10, Application US/09221013A
; Patent No. 6495740
; GENERAL INFORMATION:
; APPLICANT: Arioli, Antonio
; APPLICANT: Williamson, Richard E.
; APPLICANT: Betzner, Andreas S.
; APPLICANT: Peng, Liangcai
; TITLE OF INVENTION: Manipulation of cellulose and/or Beta-1,4-glucan
; FILE REFERENCE: 96-98
; CURRENT APPLICATION NUMBER: US/09/221,013A
; CURRENT FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: PCT/AU97/00402
; PRIOR FILING DATE: 1997-06-24
; PRIOR APPLICATION NUMBER: AU PO0699
; PRIOR FILING DATE: 1996-06-27
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 1065
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-221-013A-10

Query Match 3.6%; Score 35; DB 4; Length 1065;
Best Local Similarity 100.0%; Pred. No. 7.6e-27;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 270 TANTVLSILAVDYPVDKVSCYVSDDGAAMLSFESL 304

```
                    ||:|||||||||||||||||||:|,|||||:|||||||
Db        355 TANTVLSILAVDYPVDKVSCYVSDDGAAMLSFESL 389


RESULT 7
US-09-221-013A-8
; Sequence 8, Application US/09221013A
; Patent No. 6495740
; GENERAL INFORMATION:
;  APPLICANT:  Arioli, Antonio
;  APPLICANT:  Williamson, Richard E.
;  APPLICANT:  Betzner, Andreas S.
;  APPLICANT:  Peng, Liangcai
;  TITLE OF INVENTION:  Manipulation of cellulose and/or Beta-1,4-glucan
;  FILE REFERENCE:  96-98
;  CURRENT APPLICATION NUMBER: US/09/221,013A
;  CURRENT FILING DATE:  1998-12-23
;  PRIOR APPLICATION NUMBER:  PCT/AU97/00402
;  PRIOR FILING DATE:  1997-06-24
;  PRIOR APPLICATION NUMBER:  AU PO0699
;  PRIOR FILING DATE:  1996-06-27
;  NUMBER OF SEQ ID NOS:  37
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 8
;   LENGTH: 1084
;   TYPE: PRT
;   ORGANISM: Arabidopsis thaliana
US-09-221-013A-8

  Query Match             3.3%;  Score 32;  DB 4;  Length 1084;
  Best Local Similarity   100.0%;  Pred. No. 9.3e-24;
  Matches   32;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        256 FVSTVDPLKEPPLITANTVLSILAVDYPVDKV 287
              ||||||||||||||||||||||||||||||||
Db        359 FVSTVDPLKEPPLITANTVLSILAVDYPVDKV 390


RESULT 8
US-08-960-048-8
; Sequence 8, Application US/08960048
; Patent No. 6271443
; GENERAL INFORMATION:
;  APPLICANT:  Stalker, D. et al.
;  TITLE OF INVENTION:  Plant Cellulose Synthase and Promoter
;  TITLE OF INVENTION:  Sequences
;  FILE REFERENCE:  15621/01/US
;  CURRENT APPLICATION NUMBER: US/08/960,048
;  CURRENT FILING DATE:  1997-10-29
;  PRIOR APPLICATION NUMBER:  60/029,987
;  PRIOR FILING DATE:  1996-10-29
;  NUMBER OF SEQ ID NOS:  12
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 8
;   LENGTH: 881
;   TYPE: PRT
;   ORGANISM: Oryzae sativa
US-08-960-048-8

  Query Match             3.1%;  Score 30;  DB 3;  Length 881;
  Best Local Similarity   100.0%;  Pred. No. 8.2e-22;
  Matches   30;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        270 TANTVLSILAVDYPVDKVSCYVSDDGAAML 299
              ||||||||||||||||||||||||||||||
Db        172 TANTVLSILAVDYPVDKVSCYVSDDGAAML 201


RESULT 9
US-09-838-586-8
; Sequence 8, Application US/09838586
; Patent No. 6576818
; GENERAL INFORMATION:
;  APPLICANT: Stalker, D. et al.
;  TITLE OF INVENTION: Plant Cellulose Synthase and Promoter
;  TITLE OF INVENTION:  Sequences
;  FILE REFERENCE: 15621/02/US
;  CURRENT APPLICATION NUMBER: US/09/838,586
;  CURRENT FILING DATE:  2001-04-18
;  PRIOR APPLICATION NUMBER: 60/029,987
;  PRIOR FILING DATE: 1996-10-29
;  PRIOR APPLICATION NUMBER: 08/960,048
;  PRIOR FILING DATE: 1997-10-29
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8
;   LENGTH: 881
;   TYPE: PRT
;   ORGANISM: Oryzae sativa
US-09-838-586-8

  Query Match             3.1%;  Score 30;  DB 4;  Length 881;
  Best Local Similarity   100.0%;  Pred. No. 8.7e-22;
  Matches   30;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        270 TANTVLSILAVDYPVDKVSCYVSDDGAAML 299
              ||||||||||||||||||||||||||||||
Db        172 TANTVLSILAVDYPVDKVSCYVSDDGAAML 201


RESULT 10
US-09-221-013A-6
; Sequence 6, Application US/09221013A
; Patent No. 6495740
; GENERAL INFORMATION:
;  APPLICANT:  Arioli, Antonio
;  APPLICANT:  Williamson, Richard E.
;  APPLICANT:  Betzner, Andreas S.
;  APPLICANT:  Peng, Liangcai
;  TITLE OF INVENTION:  Manipulation of cellulose and/or Beta-1,4-glucan
;  FILE REFERENCE:  96-98
;  CURRENT APPLICATION NUMBER: US/09/221,013A
;  CURRENT FILING DATE:  1998-12-23
;  PRIOR APPLICATION NUMBER:  PCT/AU97/00402
;  PRIOR FILING DATE:  1997-06-24
;  PRIOR APPLICATION NUMBER:  AU PO0699
;  PRIOR FILING DATE:  1996-06-27
;  NUMBER OF SEQ ID NOS:  37
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 6
;   LENGTH: 1081
;   TYPE: PRT
;   ORGANISM: Arabidopsis thaliana
US-09-221-013A-6

  Query Match             2.4%;  Score 23;  DB 4;  Length 1081;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-14;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        376 PEEGWTMQDGTPWPGNNTRDHPG 398
              |||||||||||||||||||||||
Db        477 PEEGWTMQDGTPWPGNNTRDHPG 499


RESULT 11
US-09-221-013A-12
; Sequence 12, Application US/09221013A
; Patent No. 6495740
; GENERAL INFORMATION:
;  APPLICANT:  Arioli, Antonio
;  APPLICANT:  Williamson, Richard E.
;  APPLICANT:  Betzner, Andreas S.
;  APPLICANT:  Peng, Liangcai
;  TITLE OF INVENTION:  Manipulation of cellulose and/or Beta-1,4-glucan
;  FILE REFERENCE:  96-98
;  CURRENT APPLICATION NUMBER: US/09/221,013A
```

```
 CURRENT FILING DATE:  1998-12-23
 PRIOR APPLICATION NUMBER:  PCT/AU97/00402
 PRIOR FILING DATE:  1997-06-24
 PRIOR APPLICATION NUMBER:  AU PO0699
 PRIOR FILING DATE:  1996-06-27
 NUMBER OF SEQ ID NOS:  37
 SOFTWARE:  PatentIn Ver. 2.0
SEQ ID NO 12
  LENGTH: 1081
  TYPE: PRT
  ORGANISM: Arabidopsis thaliana
-09-221-013A-12

Query Match              2.4%;  Score 23;  DB 4;  Length 1081;
Best Local Similarity   100.0%;  Pred. No. 1.6e-14;
Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

        376 PEEGWTMQDGTPWPGNNTRDHPG 398
            |||||||||||||||||||||||
        477 PEEGWTMQDGTPWPGNNTRDHPG 499


SULT 12
-09-221-013A-2
Sequence 2, Application US/09221013A
Patent No. 6495740
GENERAL INFORMATION:
 APPLICANT:  Arioli, Antonio
 APPLICANT:  Williamson, Richard E.
 APPLICANT:  Betzner, Andreas S.
 APPLICANT:  Peng, Liangcai
 TITLE OF INVENTION:  Manipulation of cellulose and/or Beta-1,4-glucan
 FILE REFERENCE:  96-98
 CURRENT APPLICATION NUMBER: US/09/221,013A
 CURRENT FILING DATE:  1998-12-23
 PRIOR APPLICATION NUMBER:  PCT/AU97/00402
 PRIOR FILING DATE:  1997-06-24
 PRIOR APPLICATION NUMBER:  AU PO0699
 PRIOR FILING DATE:  1996-06-27
 NUMBER OF SEQ ID NOS:  37
 SOFTWARE:  PatentIn Ver. 2.0
SEQ ID NO 2
  LENGTH: 629
  TYPE: PRT
  ORGANISM: Arabidopsis thaliana
S-09-221-013A-2

Query Match              2.1%;  Score 21;  DB 4;  Length 629;
Best Local Similarity   100.0%;  Pred. No. 1.1e-12;
Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        378 EGWTMQDGTPWPGNNTRDHPG 398
              |||||||||||||||||||||
Db         27 EGWTMQDGTPWPGNNTRDHPG 47


ESULT 13
S-09-134-001C-4266
Sequence 4266, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
 APPLICANT: Lynn Doucette-Stamm et al
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
 TITLE OF INVENTION:  EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
 FILE REFERENCE: GTC-007
 CURRENT APPLICATION NUMBER: US/09/134,001C
 CURRENT FILING DATE:  1998-08-13
 PRIOR APPLICATION NUMBER: US 60/064,964
 PRIOR FILING DATE: 1997-11-08
 PRIOR APPLICATION NUMBER: US 60/055,779
 PRIOR FILING DATE: 1997-08-14
 NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 4266
;   LENGTH: 725
;   TYPE: PRT
;   ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4266

  Query Match              1.8%;  Score 8;  DB 4;  Length 728;
  Best Local Similarity   100.0%;  Pred. No. 28;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        644 PFIKEAIQ 651
              ||||||||
Db        273 PFIKEAIQ 280


RESULT 14
US-09-107-532A-6959
; Sequence 6959, Application US/09107532A
; Patent No. 6583275
;  GENERAL INFORMATION:
;       APPLICANT: Lynn A [illegible] and David Bush
;       TITLE OF INVENTION: [illegible] AND AMINO ACID SEQUENCES RELATING TO
;                           [illegible] FOR DIAGNOSTICS AND THERAPEUTICS
;       NUMBER OF SEQUENCES: [illegible]
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: GENEMED THERAPEUTICS CORPORATION
;            STREET: 100 Beaver Street
;            CITY: Waltham
;            STATE: Massachusetts
;            COUNTRY: USA
;            ZIP: 02354
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: CD-ROM IBM-PC
;            COMPUTER: PC
;            OPERATING SYSTEM: <Unknown>
;            SOFTWARE: ASCII
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/107,532A
;            FILING DATE: 30-Jun-1998
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 60/046,598
;            FILING DATE: 14 May 1998
;            APPLICATION NUMBER: 60/051571
;            FILING DATE: July 2, 1997
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Ariniello, Pamela Deneke
;            REGISTRATION NUMBER: 40,488
;            REFERENCE/DOCKET NUMBER: GTC-012
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (781)893-5007
;            TELEFAX: (781)893-8277
;   INFORMATION FOR SEQ ID NO: 6959:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 76 amino acids
;            TYPE: amino acid
;            TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       HYPOTHETICAL: YES
;       ORIGINAL SOURCE:
;            ORGANISM: Enterococcus faecium
;       FEATURE:
;            NAME/KEY: misc_feature
;            LOCATION: (B) LOCATION 1...76
;       SEQUENCE DESCRIPTION: SEQ ID NO: 6959:
US-09-107-532A-6959

  Query Match              0.7%;  Score 7;  DB 4;  Length 76;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        448 SAVLTNA 454
              |||||||
Db          2 SAVLTNA 8
```

RESULT 15
US-09-134-001C-4274
; Sequence 4274, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4274
; LENGTH: 124
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4274

Query Match 0.7%; Score 7; DB 4; Length 124;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 647 KEAIQVI 653
||||||| |
Db 3 KEAIQVI 9

Search completed: October 29, 2003, 18:37:31
Job time : 30.3752 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 29, 2003, 18:29:38 ; Search time 75.6673 Seconds
(without alignments)
2051.541 Million cell updates/sec

Title: US-09-980-043-2
Perfect score: 978
Sequence: 1 MMESGAPICHTCGEQVGHDA.....PFVNKVDNTLVAETCISIDC 978

Scoring table: CLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1107863 seqs, 158726573 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post processing: Listing first 45 summaries

Database : A_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	722	73.8	946	22	AAB37893	Populus tremuloide
2	61	6.2	974	19	AAW73308	Cellulose synthase
3	61	6.2	974	19	AAW60039	Cotton cellulose s
4	47	4.8	958	23	ABB92981	Herbicidally activ
5	44	4.5	56	22	AAB37894	Populus tremuloide
6	43	4.4	189	21	AAG50463	Arabidopsis thalia
7	43	4.4	277	21	AAG06365	Arabidopsis thalia
8	43	4.4	285	21	AAG50462	Arabidopsis thalia
9	43	4.4	288	21	AAG06364	Arabidopsis thalia
10	43	4.4	289	21	AAG06363	Arabidopsis thalia
11	43	4.4	296	21	AAG50461	Arabidopsis thalia
12	43	4.4	955	21	AAB16338	Eucalyptus grandis
13	40	4.1	239	21	AAB16279	Eucalyptus grandis
14	40	4.1	547	19	AAW46202	Rice RSW1-like cel
15	40	4.1	1074	21	AAY84110	Amino acid sequenc
16	40	4.1	1074	21	AAY84112	Amino acid sequenc
17	40	4.1	1074	21	AAY84117	Amino acid sequenc
18	39	4.0	469	21	AAB16281	Eucalyptus grandis
19	39	4.0	686	19	AAW60040	Cotton cellulose s
20	39	4.0	1039	19	AAW73309	Cellulose synthase
21	39	4.0	1043	23	ABB93701	Herbicidally activ
22	36	3.7	793	21	AAY58839	Soybean cellulose
23	35	3.6	959	21	AAG48735	Arabidopsis thalia
24	35	3.6	1054	21	AAG48734	Arabidopsis thalia
25	35	3.6	1065	19	AAW33819	Arabidopsis cellul
26	35	3.6	1065	21	AAG48733	Arabidopsis thalia
27	35	3.6	1065	23	ABB93356	Herbicidally activ
28	32	3.3	1080	22	AAB37895	Arabidopsis thalia
29	32	3.3	1084	19	AAW33818	Arabidopsis cellul
30	32	3.3	1084	23	ABB93304	Herbicidally activ
31	32	3.3	1084	23	ABB93949	Herbicidally activ
32	30	3.1	196	21	AAB16329	Eucalyptus grandis
33	30	3.1	320	21	AAY58833	Corn cellulose syn
34	30	3.1	534	21	AAB16276	Eucalyptus grandis
35	30	3.1	600	21	AAB16304	Eucalyptus grandis
36	30	3.1	1043	21	AAY84119	Amino acid sequenc
37	30	3.1	1076	21	AAY84121	DNA encoding a mai
38	30	3.1	1077	21	AAY84108	Amino acid sequenc
39	30	3.1	1077	21	AAY84114	Amino acid sequenc
40	30	3.1	1079	23	AAU10496	Corn cellulose syn
41	29	3.0	1069	23	ABB93411	Herbicidally activ
42	29	3.0	1088	23	ABB91831	Herbicidally activ
43	28	2.9	127	21	AAB16275	Pinus radiata cell
44	26	2.7	206	21	AAB16278	Pinus radiata cell
45	24	2.5	821	23	AAU10495	Corn cellulose syn

ALIGNMENTS

RESULT 1
AAB37893
ID AAB37893 standard; Protein; 946 AA.
XX
AC AAB37893;
XX
DT 07-MAR-2001 (first entry)
XX
DE Populus tremuloides cellulose synthase CelA.
XX
KW Aspen; CelA; cellulose synthase; arabidopsis; wood;
KW mechanical stress response element; MSRE; cellulose biosynthesis.
XX
OS Populus tremuloides.
XX
PN WO200071670-A2.
XX
PD 30-NOV-2000.
XX
PF 18-MAY-2000; 2000WO-US13637.
XX
PR 21-MAY-1999; 99US-0135280.
XX
PA (UNMT) UNIV MICHIGAN TECHNOLOGICAL.
XX
PI Chiang VL, Wu L, Chandrashekhar JP, Carraway DT;
XX
DR WPI; 2001-032025/04.
DR N-PSDB; AAC65448.
XX
PT Novel polynucleotide encoding cellulose synthase useful for enhancing
PT cellulose biosynthesis, modifying lignin biosynthesis, and for

accelerating growth of plants and for upregulating cellulose synthase levels -

Claim 8; Fig 1; 65pp; English.

The present sequence is given in a specification relating to polynucleotides encoding cellulose synthase enzymes from Populus tremuloides and Arabidopsis. The cellulose synthases are designated PtCelA and AraxCelA respectively. The polynucleotides may be introduced into plants where they are useful for altering the growth of a plant. Vectors comprising a cellulose synthase promoter, or a fragment of the promoter containing one or more mechanical stress response elements (MSRE), operably linked to a cellulose synthase gene are useful for causing stress-induced gene expression in a plant cell. A cDNA encoding a protein that binds to a positive MSRE may be delivered into the plant to increase expression of cellulose synthase. A cDNA in an antisense orientation may be delivered to reduce expression of cellulose synthase. The polynucleotides enable the development of transgenic tree varieties having increased cellulose content, decreased lignin content and therefore improved wood fibre characteristics. The cellulose synthase promoters are useful for synthesising greater quantities of high-crystalline cellulose in plant and preferably in trees, which permits an increase in proportion of cellulose in transgenic plants, greater strength of juvenile wood and fiber and acceleration of overall growth rate.

Sequence 946 AA;

```
Query Match             73.8%;  Score 722;  DB 22;  Length 946;
Best Local Similarity   100.0%;  Pred. No. 0;
Matches  722;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

        257 VSTVDPLKEPPLITANTVLSILAVDYPVDKVSCYVSDDGAAMLSFESLVETAEFARKWVP 316
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
        225 VSTVDPLKEPPLITANTVLSILAVDYPVDKVSCYVSDDGAAMLSFESLVETAEFARKWVP 284

        317 FCKKFSIEPRAPEFYFSQKIDYLKDKVQPSFVKERRAMKRDYEEYKVRVNALVAKAQKTP 376
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
        285 FCKKFSIEPRAPEFYFSQKIDYLKDKVQPSFVKERRAMKRDYEEYKVRVNALVAKAQKTP 344

        377 EEGWTMQDGTPWPGNNTRDHPGHDSGLPWEILGARDIEGNELPRLVYVSREKRPGYQHHK 436
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
        345 EEGWTMQDGTPWPGNNTRDHPGHDSGLPWEILGARDIEGNELPRLVYVSREKRPGYQHHK 404

        437 KAGAENALVRVSAVLTNAPYILNVDCDHYVNNSKAVREAMCILMDPQVGRDVCYVQFPQR 496
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
        405 KAGAENALVRVSAVLTNAPYILNVDCDHYVNNSKAVREAMCILMDPQVGRDVCYVQFPQR 464

        497 FDGIDKSDRYANRNVVFFDVNMKGLDGIQGPVYVGTGCVFNRQALYGYGPPSMPSLRKRK 556
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
        465 FDGIDKSDRYANRNVVFFDVNMKGLDGIQGPVYVGTGCVFNRQALYGYGPPSMPSLRKRK 524

        557 DSSSCFSCCCPSKKKPAQDPAEVYRDAKREDLNAAIFNLTEIDNYDEHERSMLISQLSFE 616
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
        525 DSSSCFSCCCPSKKKPAQDPAEVYRDAKREDLNAAIFNLTEIDNYDEHERSMLISQLSFE 584

        617 KTFGLSSVFIESTLMENGGVPESANSPPFIKEAIQVIGCGYEEKTEWGKQIGWIYGSVTE 676
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
        585 KTFGLSSVFIESTLMENGGVPESANSPPFIKEAIQVIGCGYEEKTEWGKQIGWIYGSVTE 644

        677 DILSGFKMHCRGWRSIYCMPVRPAFKGSAPINLSDRLHQVLRWALGSVEIFFSRHCPLWY 736
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
        645 DILSGFKMHCRGWRSIYCMPVRPAFKGSAPINLSDRLHQVLRWALGSVEIFFSRHCPLWY 704

        737 GFGGGRLKWLQRLAYINTIVYPFTSLPLIAYCTIPAVCLLTGKFIIPTLSNLASMLFLGL 796
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
        705 GFGGGRLKWLQRLAYINTIVYPFTSLPLIAYCTIPAVCLLTGKFIIPTLSNLASMLFLGL 764

        797 FISIIVTAVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQGFLKMLAGIDTNFTVTA 856
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
        765 FISIIVTAVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQGFLKMLAGIDTNFTVTA 824

        857 KAAEDAEFGELYMVKWTTLLIPPTTLLIINMSGCAGFSDALNKGYEAWGPLFGKVFFAFW 916
```

```
Db      825 KAAEDAEFGELYMVKWTTLLIPPTTLLIINMSGCAGFSDALNKGYEAWGPLFGKVFFAFW 884

Qy      917 VILHLYPFLKGLMGRQNLTPTIVVLWSVLLASVFSLVWVKINPFVNKVDNTLVAETCISI 976
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      885 VILHLYPFLKGLMGRQNLTPTIVVLWSVLLASVFSLVWVKINPFVNKVDNTLVAETCISI 944

Qy      977 DC 978
            ||
Db      945 DC 946
```

RESULT 2
AAW73308
ID AAW73308 standard; Protein; 974 AA
XX
AC AAW73308;
XX
DT 08-FEB-1999 (first entry)
XX
DE Cellulose synthase PtsA
XX
KW Cellulose synthase; [illegible]
XX
OS Gossypium hirsutum
XX
PN EP875579-A2.
XX
PD 04-NOV-1998.
XX
PF 31-MAR-1998; 98EP-030489
XX
PR 01-APR-1997; 97JP-008[illegible]
XX
PA (NISN) NISSHINBO IND INC.
XX
PI Aotsuka S, Hasegawa O, Hayashi T, Ihara Y;
XX
DR WPI; 1998-559440/48.
DR N-PSDB; AAV08372.
XX
PT New DNA encoding cotton cellulose synthase proteins - useful for
PT controlling cellulose synthesis in prokaryotic or eukaryotic cells
XX
PS Claim 1; Page 14-18; 51pp; English.
XX
CC This sequence is a cellulose synthase of the invention. The DNA can
CC be used for controlling cellulose synthesis in a cell by introducing the
CC DNA into the cell, expressing RNA with a nucleotide sequence homologous
CC to the DNA or a nucleotide sequence complementary to the DNA. The cells
CC are prokaryotic cells i.e. acetobacterium and/or eukaryotic cells such as
CC yeasts i.e. Saccharomyces spp., plant cells (cotton plants) and cultured
CC cells of mammals etc.
XX
SQ Sequence 974 AA;

```
  Query Match             6.2%;  Score 61;  DB 19;  Length 974;
  Best Local Similarity   [illegible]0%;  Pred. No. 6.5e-54;
  Matches   61;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        799 SIIVTAVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQGFLKMLAGIDTNFTVTAKA 858
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        794 SIIVTAVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQGFLKMLAGIDTNFTVTAKA 853

Qy        859 A 859
              |
Db        854 A 854
```

RESULT 3
AAW60039
ID AAW60039 standard; Protein; 974 AA.
XX

```
AC   AAW60039;
XX
DT   28-SEP-1998  (first entry)
XX
DE   Cotton cellulose synthase CelA1.
XX
KW   Cellulose synthase; CelA1; cotton; forest tree; plant fibre;
KW   wood; transgenic plant; ss.
XX
OS   Gossypium hirsutum Acala SJ 2.
XX
FH   Key             Location/Qualifiers
FT   Region          175..295
FT                   /label= H-1
FT   Region          280..295
FT                   /label= U-1
FT   Region          296..420
FT                   /label= P-CR
FT   Region          421..539
FT                   /label= H-2
FT   Region          450..467
FT                   /label= U-2
FT   Region          540..658
FT                   /label= HVR
FT   Region          659..880
FT                   /label= H-3
FT   Region          667..688
FT                   /label= U-3
FT   Region          706..722
FT                   /label= U-4
FT   Modified-site   81
FT                   /note= "N-glycosylated"
FT   Modified-site   466
FT                   /note= "N-glycosylated"
FT   Modified-site   703
FT                   /note= "N-glycosylated"
FT   Modified-site   846
FT                   /note= "N-glycosylated"
FT   Misc difference 293
FT                   /note= "critical Asp residue"
FT   Misc difference 461
FT                   /note= "critical Asp residue"
FT   Misc-difference 672
FT                   /note= "critical Asp residue"
FT   Peptide         710..714
FT                   /note= "QXXRW motif"
XX
PN   WO9818949 A2
XX
PD   07-MAY-1998
XX
PF   29-OCT-1997;   97WO-US19529.
XX
PR   29-OCT-1996;   96US-0029987.
XX
PA   (CALG ) CALGENE INC.
XX
PI   Delmer DP,  Pear JR,  Stalker DM;
XX
DR   WPI; 1998-272238/24.
DR   N-PSDB; AAV34432.
XX
PT   New isolated plant cellulose synthesis enzyme DNA  is used to
PT   develop products for producing modified plant fibre products,
PT   particularly for modifying cotton fibre and wood quality
XX
PS   Example 5; Fig 3A-C; 49pp; English.
XX
CC   This full-length polypeptide comprises cotton cellulose synthase
CC   CelA1.  Its amino acid sequence was deduced from a cDNA clone (see
CC   AAV60039) isolated from a cotton fibre cDNA library.  Expression of
CC   CelA1 and CelA2 (see AAW60040) in developing fibre is approximately
CC   500 times higher than the level of expression in other cotton
CC   tissues; approximately 1-2% of the 24 dpa mRNA is CelA mRNA.
CC   Sequence analysis of CelA1 and CelA2 showed regions of high homology
CC   (H-1, H-2 and H-3) between plant and bacteral proteins, as well as
CC   a region (P-CR) that is conserved among plant CelA genes, and
CC   another region (HVR) that is hypervariable between plant genes.
CC   Regions U-1 to U-4 are predicted to be critical for UDP-glc binding
CC   and catalysis in bacterial CelA proteins.  The isolated CelA cDNA
CC   clones are used in claimed methods for modifying fibre phenotype
CC   in a cotton plant and for modifying wood quality phenotype in a
CC   forest tree species.  Lignin can be reduced by over-expressing the
CC   cellulose synthase product and diverting production into cellulose.
XX
SQ   Sequence   974 AA;

  Query Match             6.2%;  Score 61;  DB 19;  Length 974;
  Best Local Similarity   100.0%;  Pred. No. 6.5e-54;
  Matches   61;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        799 SIIVTAVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQGFLKMLAGIDTNFTVTAKA 858
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        794 SIIVTAVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQGFLKMLAGIDTNFTVTAKA 853

Qy        859 A 859
              |
Db        854 A 854


RESULT 4
ABB92981
ID   ABB92981 standard; Protein; 958 AA.
XX
AC   ABB92981;
XX
DT   31-MAY-2002  (first entry)
XX
DE   Herbicidally active polypeptide SEQ ID NO 2192.
XX
KW   Herbicidal; plant; agriculture; herbicide.
XX
OS   Arabidopsis thaliana.
XX
PN   WO200210210-A2.
XX
PD   07-FEB-2002.
XX
PF   28-AUG-2001; 2001WO-EP09892.
XX
PR   28-AUG-2001; 2001WO-EP09892.
XX
PA   (FARB ) BAYER AG.
XX
PI   Tietjen K,  Weidler M;
XX
DR   WPI; 2002-269010/31.
XX
PT   Identifying plant target proteins for herbicidally active compounds,
PT   comprising aligning and comparing nucleic acid or amino acid sequences
PT   from plant with nucleic acid or amino acid sequences from non-plant
PT   organisms -
XX
PS   Claim 5; SEQ ID NO 2192; 261pp + Sequence Listing; English.
XX
CC   The invention relates to identifying target proteins
CC   (ABB90790-ABB94016) for herbicidally active compounds, comprising
CC   aligning and comparing nucleic acid or amino acid sequences from plant
CC   with nucleic acid or amino acid sequences from non-plant organisms using
CC   suitable search parameters, where plant sequences having an E-value
CC   greater by a factor of 3 than the E-value of most similar non-plant
CC   sequences are selected. The polypeptides or nucleic acids encoding them
CC   are useful for identifying modulators. The identified modulators are
CC   useful as herbicides.
XX
SQ   Sequence   958 AA;
```

```
 Query Match              4.8%;  Score 47;  DB 23;  Length 958;
 Best Local Similarity   100.0%;  Pred. No. 2.6e-39;
 Matches   47;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

/        409 GARDIEGNELPRLVYVSREKRPGYQHHKKAGAENALVRVSAVLTNAP 455
            |||||||||||||||||||||||||||||||||||||||||||||||
D        408 GARDIEGNELPRLVYVSREKRPGYQHHKKAGAENALVRVSAVLTNAP 454


ESULT 5
AB37894
D   AAB37894 standard; Protein; 56 AA.
X
C   AAB37894;
X
T   07-MAR-2001  (first entry)
X
E   Populus tremuloides CelA N-terminal sequence.
X
W   Aspen; CelA; cellulose synthase; arabidopsis; wood;
W   mechanical stress response element; MSRE; cellulose biosynthesis
X
S   Populus tremuloides.
X
N   WO200071670-A2.
X
D   30-NOV-2000.
X
F   18-MAY-2000; 2000WO-US13637.
X
R   21-MAY-1999;   99US-0135280.
X
A   (UNMT ) UNIV MICHIGAN TECHNOLOGICAL.
X
I   Chiang VL,  Wu L,  Chandrashekhar JP,  Carraway DT;
X
R   WPI; 2001-032025/04.
R   N-PSDB; AAC65449.
X
T   Novel polynucleotide encoding cellulose synthase useful for enhancing
T   cellulose biosynthesis, modifying lignin biosynthesis, and for
PT  accelerating growth of plants and for upregulating cellulose synthase
PT  levels -
XX
PS  Claim 1; Fig 4; 65pp; English.
XX
CC  The present sequence is given in a specification relating to
CC  polynucleotides encoding cellulose synthase enzymes from Populus
CC  tremuloides and Arabidopsis. The cellulose synthases are designated
CC  PtCelA and AraxCelA respectively. The polynucleotides may be introduced
CC  into plants where they are useful for altering the growth of a plant.
CC  Vectors comprising a cellulose synthase promoter, or a fragment of the
CC  promoter containing one or more mechanical stress response elements
CC  (MSRE), operably linked to a cellulose synthase gene are useful for
CC  causing stress-induced gene expression in a plant cell. A cDNA encoding a
CC  protein that binds to a positive MSRE may be delivered into the plant to
CC  increase expression of cellulose synthase. A cDNA in an antisense
CC  orientation may be delivered to reduce expression of cellulose synthase.
CC  The polynucleotides enable the development of transgenic tree varieties
CC  having increased cellulose content, decreased lignin content and
CC  therefore improved wood fibre characteristics. The cellulose synthase
CC  promoters are useful for synthesising greater quantities of
CC  high-crystalline cellulose in plant and preferably in trees, which
CC  permits an increase in proportion of cellulose in transgenic plants,
CC  greater strength of juvenile wood and fiber and acceleration of overall
CC  growth rate.
CC  Note: This polypeptide sequence is given in the specification as
CC  SEQ ID NO: 4. SEQ ID NO: 4 is referred to in claim 1 but it is described
CC  as a nucleotide sequence.
XX
SQ  Sequence   56 AA;
```

```
  Query Match              [illegible]%;  Score 44;  DB [illegible];  Length 56;
  Best Local Similarity    [illegible]%;  Pred. No. [illegible];
  Matches   44;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MMESGAPICHTCGE[illegible]GHDANGELFVACHECSYPMCKSCFEFEI 44
              |||| ||| .||| .  ||.|'||  .||| ||||. ||||.||
Db          1 MMESGAPICHTCGE[illegible]GHDANGELFVACHECSYPMCKSCFEFEI 44


RESULT 6
AAG50463
ID   AAG50463 standard; P[illegible] 168 AA.
XX
AC   AAG50463;
XX
DT   18-OCT-2000  (first [illegible])
XX
DE   Arabidopsis thaliana [illegible] fragment SEQ ID NO: 63954.
XX
KW   Protein identification; [illegible] pathway; metabolic pathway;
KW   hybridisation assay; [illegible]; [illegible] expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-[illegible]
XX
PR   25-FEB-1999;   99US-[illegible]825.
PR   05-MAR-1999;   99US-[illegible]180.
PR   09-MAR-1999;   99US-[illegible]648.
PR   23-MAR-1999;   99US-[illegible]788.
PR   25-MAR-1999;   99US-[illegible]764.
PR   29-MAR-1999;   99US-[illegible]785.
PR   01-APR-1999;   99US-[illegible]462.
PR   06-APR-1999;   99US-[illegible]234.
PR   08-APR-1999;   99US-[illegible]714.
PR   16-APR-1999;   99US-[illegible]845.
PR   19-APR-1999;   99US-[illegible]079.
PR   21-APR-1999;   99US-[illegible]449.
PR   23-APR-1999;   99US-[illegible]510.
PR   23-APR-1999;   99US-[illegible]691.
PR   28-APR-1999;   99US-[illegible]449.
PR   30-APR-1999;   99US-[illegible]048.
PR   30-APR-1999;   99US-[illegible]409.
PR   04-MAY-1999;   99US-[illegible]484.
PR   05-MAY-1999;   99US-[illegible]485.
PR   06-MAY-1999;   99US-[illegible]486.
PR   06-MAY-1999;   99US-[illegible]487.
PR   07-MAY-1999;   99US-[illegible]863.
PR   11-MAY-1999;   99US-[illegible]256.
PR   14-MAY-1999;   99US-[illegible]218.
PR   14-MAY-1999;   99US-[illegible]219.
PR   14-MAY-1999;   99US-[illegible]221.
PR   14-MAY-1999;   99US-[illegible]370.
PR   18-MAY-1999;   99US-[illegible]768.
PR   19-MAY-1999;   99US-[illegible]941.
PR   20-MAY-1999;   99US-[illegible]124.
PR   21-MAY-1999;   99US-[illegible]353.
PR   24-MAY-1999;   99US-[illegible]629.
PR   25-MAY-1999;   99US-[illegible]021.
PR   27-MAY-1999;   99US-[illegible]392.
PR   28-MAY-1999;   99US-[illegible]782.
PR   01-JUN-1999;   99US-[illegible]222.
PR   03-JUN-1999;   99US-[illegible]528.
PR   04-JUN-1999;   99US-[illegible]502.
PR   07-JUN-1999;   99US-[illegible]724.
PR   08-JUN-1999;   99US-[illegible]094.
PR   10-JUN-1999;   99US-[illegible]540.
PR   10-JUN-1999;   99US-[illegible]847.
```

PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462
PR 18 JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999, 99US-0139763.
PR 21 JUN-1999; 99US-0139817
PR 22-JUN-1999; 99US-0139899
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991
PR 30 JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL 1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999, 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333
PR 19-JUL-1999; 99US-0144334.
PR 19 JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20 JUL-1999; 99US-0144884.
PR 21 JUL-1999, 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL 1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087
PR 22 JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL 1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10 AUG 1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 4.4%; Score 43; DB 21; Length 189;
Best Local Similarity 100.0%; Pred. No. 8.6e-36;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 805 VLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQGFLKMLAG 847
Db 15 VLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQGFLKMLAG 57

SULT 7
G06365
AAG06365 standard; Protein; 277 AA.

AAG06365;

17-OCT-2000 (first entry)

Arabidopsis thaliana protein fragment SEQ ID NO: 3112.

Protein identification; signal transduction pathway; metabolic pathway; hybridisation assay; genetic mapping; gene expression control; promoter; termination sequence.

Arabidopsis thaliana.

EP1033405-A2.

06-SEP-2000.

25-FEB-2000; 2000EP-0301439.

25-FEB-1999; 99US-0121825.
05-MAR-1999; 99US-0123180.
09-MAR-1999; 99US-0123548.
23-MAR-1999; 99US-0125788.
25-MAR-1999; 99US-0126264.
29-MAR-1999; 99US-0126785.
01-APR-1999; 99US-0127462.
06-APR-1999; 99US-0128234.
08-APR-1999; 99US-0128714.
16-APR-1999; 99US-0129845.
19-APR-1999; 99US-0130077.
21-APR-1999; 99US-0130449.
23-APR-1999; 99US-0130510.
23-APR-1999; 99US-0130891.
28-APR-1999; 99US-0131449.
30-APR-1999; 99US-0132048.
30-APR-1999; 99US-0132407.
04-MAY-1999; 99US-0132484.
05-MAY-1999; 99US-0132485.
06-MAY-1999; 99US-0132486.
06-MAY-1999; 99US-0132487.
07-MAY-1999; 99US-0132863.
11-MAY-1999; 99US-0134256.
14-MAY-1999; 99US-0134218.
14-MAY-1999; 99US-0134219.
14-MAY-1999; 99US-0134221.
14-MAY-1999; 99US-0134370.
18-MAY-1999; 99US-0134768.
19-MAY-1999; 99US-0134941.
20-MAY-1999; 99US-0135124.
21-MAY-1999; 99US-0135353.
24-MAY-1999; 99US-0135629.
25-MAY-1999; 99US-0136021.
27-MAY-1999; 99US-0136392.
28-MAY-1999; 99US-0136782.
01-JUN-1999; 99US-0137222.
03-JUN-1999; 99US-0137528.
04-JUN-1999; 99US-0137502.
07-JUN-1999; 99US-0137724.
08-JUN-1999; 99US-0138094.
10-JUN-1999; 99US-0138540.
10-JUN-1999; 99US-0138847.
14-JUN-1999; 99US-0139119.
16-JUN-1999; 99US-0139452.
16-JUN-1999; 99US-0139453.
17-JUN-1999; 99US-0139492.
18-JUN-1999; 99US-0139454.
18-JUN-1999; 99US-0139455.
18-JUN-1999; 99US-0139456.
18-JUN-1999; 99US-0139457.
18-JUN-1999; 99US-0139458.
18-JUN-1999; 99US-0139459.
18-JUN-1999; 99US-0139460.
18-JUN-1999; 99US-0139461.
18-JUN-1999; 99US-0139462.
18-JUN-1999; 99US-0139463.
18-JUN-1999; 99US-0139750.
18-JUN-1999; 99US-0139763.
21-JUN-1999; 99US-0139817.
22-JUN-1999; 99US-0139899
23-JUN-1999; 99US-0140353.
23-JUN-1999; 99US-0140354.
24-JUN-1999; 99US-0140695.
28-JUN-1999; 99US-0140823.
29-JUN-1999; 99US-0140991.
30-JUN-1999; 99US-0141287
01-JUL-1999; 99US-0141842.
01-JUL-1999; 99US-0141854
02-JUL-1999; 99US-0141859
06-JUL-1999; 99US-0142447.
08-JUL-1999; 99US-0142804.
09-JUL-1999; 99US-0142965.
12-JUL-1999; 99US-0142977.
13-JUL-1999; 99US-0143542.
14-JUL-1999; 99US-0143624.
15-JUL-1999; 99US-0144005.
16-JUL-1999; 99US-0144085.
16-JUL-1999; 99US-0144086.
19-JUL-1999; 99US-0144325.
19-JUL-1999; 99US-0144331
19-JUL-1999; 99US-0144332.
19-JUL-1999; 99US-0144333.
19-JUL-1999; 99US-0144334.
19-JUL-1999; 99US-0144335.
20-JUL-1999; 99US-0144352.
20-JUL-1999; 99US-0144632.
20-JUL-1999; 99US-0144884
21-JUL-1999; 99US-0144814.
21-JUL-1999; 99US 0145086.
21-JUL-1999; 99US-0145088.
22-JUL-1999; 99US-0145085.
22-JUL-1999; 99US-0145087.
22-JUL 1999; 99US-0145089.
22-JUL 1999; 99US-0145192.
23-JUL-1999; 99US-0145145.
23-JUL-1999; 99US-0145218.
23-JUL-1999; 99US-0145224.
26-JUL-1999; 99US-0145276.
27-JUL-1999; 99US-0145913.
27-JUL-1999; 99US-0145918.
27-JUL-1999; 99US-0145919.
28-JUL-1999; 99US-0145951.
02-AUG-1999; 99US-0146386.
02-AUG-1999; 99US-0146388.
02-AUG-1999; 99US-0146389
03-AUG-1999; 99US-0147038
04-AUG-1999; 99US-0147204.
04-AUG-1999; 99US-0147302.
05-AUG-1999; 99US-0147192.
05-AUG-1999; 99US-0147260.
06-AUG-1999; 99US-0147303.
06-AUG-1999; 99US-0147416.
09-AUG-1999; 99US-0147493.
09-AUG-1999; 99US-0147935.
10-AUG-1999; 99US-0148171.
11-AUG-1999; 99US-0148319.
12-AUG-1999; 99US-0148341.
13-AUG-1999; 99US-0148565.
13-AUG-1999; 99US-0148684.
16-AUG-1999; 99US-0149368.
17-AUG-1999; 99US-0149175.
18-AUG-1999; 99US-0149426.

PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 4.4%; Score 43; DB 21; Length 277;
Best Local Similarity 100.0%; Pred. No. 1.2e-35;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 805 VLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQGFLKMLAG 847
||||||||||||||||||||||||||||||||||||||||||||
Db 111 VLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQGFLKMLAG 153

RESULT 8
AAG50462
ID AAG50462 standard; Protein; 285 AA.
XX
AC AAG50462;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 63953.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.

18-JUN-1999; 99US-0139750.
18-JUN-1999; 99US-0139763.
21-JUN-1999; 99US-0139817.
22-JUN-1999; 99US-0139899.
23-JUN-1999; 99US-0140353.
23-JUN-1999; 99US-0140354.
24-JUN-1999; 99US-0140695.
28-JUN-1999; 99US-0140823.
29-JUN-1999; 99US-0140991.
30-JUN-1999; 99US-0141287.
01-JUL-1999; 99US-0141842.
01-JUL-1999; 99US-0142154.
02-JUL-1999; 99US-0142055.
06-JUL-1999; 99US-0142390.
08-JUL-1999; 99US-0142803.
09-JUL-1999; 99US-0142920.
12-JUL-1999; 99US-0142977.
13-JUL-1999; 99US-0143542.
14-JUL-1999; 99US-0143624.
15-JUL-1999; 99US-0144005.
16-JUL-1999; 99US-0144085.
16-JUL-1999; 99US-0144086.
19-JUL-1999; 99US-0144325.
19-JUL-1999; 99US-0144331.
19-JUL-1999; 99US-0144332.
19-JUL-1999; 99US-0144333.
19-JUL-1999; 99US-0144334.
19-JUL-1999; 99US-0144335.
20-JUL-1999; 99US-0144352.
20-JUL-1999; 99US-0144632.
20-JUL-1999; 99US-0144884.
21-JUL-1999; 99US-0144814.
21-JUL-1999; 99US-0145086.
21-JUL-1999; 99US-0145088.
22-JUL-1999; 99US-0145085.
22-JUL-1999; 99US-0145087.
22-JUL-1999; 99US-0145089.
22-JUL-1999; 99US-0145192.
23-JUL-1999; 99US-0145145.
23-JUL-1999; 99US-0145218.
23-JUL-1999; 99US-0145224.
26-JUL-1999; 99US-0145276.
27-JUL-1999; 99US-0145913.
27-JUL-1999; 99US-0145918.
27-JUL-1999; 99US-0145919.
28-JUL-1999; 99US-0145951.
02-AUG-1999; 99US-0146386.
02-AUG-1999; 99US-0146388.
02-AUG-1999; 99US-0146389.
03-AUG-1999; 99US-0147038.
04-AUG-1999; 99US-0147204.
04-AUG-1999; 99US-0147302.
05-AUG-1999; 99US-0147192.
05-AUG-1999; 99US-0147260.
06-AUG-1999; 99US-0147303.
06-AUG-1999; 99US-0147416.
09-AUG-1999; 99US-0147493.
09-AUG-1999; 99US-0147935.
10-AUG-1999; 99US-0148171.
11-AUG-1999; 99US-0148319.
12-AUG-1999; 99US-0148341.
13-AUG-1999; 99US-0148565.
13-AUG-1999; 99US-0148684.
16-AUG-1999; 99US-0149368.
17-AUG-1999; 99US-0149175.
18-AUG-1999; 99US-0149426.
20-AUG-1999; 99US-0149722.
20-AUG-1999; 99US-0149723.
20-AUG-1999; 99US-0149929.
23-AUG-1999; 99US-0149902.
23-AUG-1999; 99US-0149930.
25-AUG-1999; 99US-0150566.
26-AUG-1999; 99US-0150884.

PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151353.
PR 31-AUG-1999; 99US-0151438
PR 01-SEP-1999; 99US-0151930
PR 07-SEP-1999; 99US-0151963.
PR 10-SEP-1999; 99US-0152020
PR 13-SEP-1999; 99US-0152258
PR 15-SEP-1999; 99US-0152718
PR 16-SEP-1999; 99US-0152836
PR 20-SEP-1999; 99US-0153173
PR 22-SEP-1999; 99US-0153448.
PR 23-SEP-1999; 99US-0153466.
PR 24-SEP-1999; 99US-0153659.
PR 28-SEP-1999; 99US-0154458.
PR 29-SEP-1999; 99US-0154596.
PR 04-OCT-1999; 99US-0155112
PR 05-OCT-1999; 99US-0155[illegible]
PR 06-OCT-1999; 99US-0155[illegible]
PR 07-OCT-1999; 99US-0155[illegible]
PR 08-OCT-1999; 99US-0155[illegible]
PR 12-OCT-1999; 99US-0156[illegible]
PR 13-OCT-1999; 99US-0156293.
PR 13-OCT-1999; 99US-0156294.
PR 13-OCT-1999; 99US-0156295.
PR 14-OCT-1999; 99US-0156325.
PR 14-OCT-1999; 99US-0156330.
PR 14-OCT-1999; 99US-0156331
PR 14-OCT-1999; 99US-0156637.
PR 14-OCT-1999; 99US-0156638.
PR 18-OCT-1999; 99US-0159584
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 4.4%; Score 43; DB 21; Length 285;
Best Local Similarity 100.0%; Pred. No. 1.3e-35;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 805 VLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQQFLKMLAG 847
Db 111 VLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQQFLKMLAG 153

RESULT 9
AAG06364
ID AAG06364 standard; Protein; 298 AA.
XX
AC AAG06364;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 3111.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;

KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.

```
R   10-SEP-1999;   99US-0153070.
R   13-SEP-1999;   99US-0153758.
R   15-SEP-1999;   99US-0154018.
R   16-SEP-1999;   99US-0154039.
R   20-SEP-1999;   99US-0154779.
R   22-SEP-1999;   99US-0155139.
R   23-SEP-1999;   99US-0155486.
R   24-SEP-1999;   99US-0155659.
R   28-SEP-1999;   99US-0156458.
R   29-SEP-1999;   99US-0156596.
R   04-OCT-1999;   99US-0157117.
R   05-OCT-1999;   99US-0157753.
R   06-OCT-1999;   99US-0157865.
R   07-OCT-1999;   99US-0158029.
R   08-OCT-1999;   99US-0158232.
R   12-OCT-1999;   99US-0158369.
R   13-OCT-1999;   99US-0159293.
R   13-OCT-1999;   99US-0159294.
R   13-OCT-1999;   99US-0159295.
R   14-OCT-1999;   99US-0159329.
R   14-OCT-1999;   99US-0159330.
R   14-OCT-1999;   99US-0159331.
R   14-OCT-1999;   99US-0159637.
R   14-OCT-1999;   99US-0159638.
R   18-OCT-1999;   99US-0159584.
R   21-OCT-1999;   99US-0160741.
R   21-OCT-1999;   99US-0160767.
R   21-OCT-1999;   99US-0160768.
R   21-OCT-1999;   99US-0160770.
R   21-OCT-1999;   99US-0160814.
R   21-OCT-1999;   99US-0160815.
R   22-OCT-1999;   99US-0160980.
R   22-OCT-1999;   99US-0160981.
R   22-OCT-1999;   99US-0160989.
R   25-OCT-1999;   99US-0161404.
R   25-OCT-1999;   99US-0161405.
R   25-OCT-1999;   99US-0161406.
R   26-OCT-1999;   99US-0161359.
R   26-OCT-1999;   99US-0161360.
R   26-OCT-1999;   99US-0161361.
R   28-OCT-1999;   99US-0161920.
R   28-OCT-1999;   99US-0161992.
R   28-OCT-1999;   99US-0161993.
R   29-OCT-1999;   99US-0162142.

 Query Match              4.4%;  Score 43;  DB 21;  Length 288;
 Best Local Similarity   100.0%;  Pred. No. 1.3e-35;
 Matches   43;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        805 VLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQGFLKMLAG 847
              |||||||||||||||||||||||||||||||||||||||||||
Db        122 VLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQGFLKMLAG 164


RESULT 10
AAG06363
ID   AAG06363 standard; Protein; 289 AA.
XX
AC   AAG06363;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 3110.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.

XX
PF   25-FEB-2000; 2000EP-0301439
XX
PR   25-FEB-1999;   99US-0121825
PR   05-MAR-1999;   99US-0123180
PR   09-MAR-1999;   99US-0123548
PR   23-MAR-1999;   99US-0125788
PR   25-MAR-1999;   99US-0126264
PR   29-MAR-1999;   99US-0126785
PR   01-APR-1999;   99US-0127462
PR   06-APR-1999;   99US-0128234
PR   08-APR-1999;   99US-0128714
PR   16-APR-1999;   99US-0129845
PR   19-APR-1999;   99US-0130027
PR   21-APR-1999;   99US-0130449
PR   23-APR-1999;   99US-0130610
PR   23-APR-1999;   99US-0130891
PR   28-APR-1999;   99US-0131449
PR   30-APR-1999;   99US-0131746
PR   30-APR-1999;   99US-0131787
PR   04-MAY-1999;   99US-0132184
PR   05-MAY-1999;   99US-0132436
PR   06-MAY-1999;   99US-0132494
PR   06-MAY-1999;   99US-0132687
PR   07-MAY-1999;   99US-0133863
PR   11-MAY-1999;   99US-0134256
PR   14-MAY-1999;   99US-0134218
PR   14-MAY-1999;   99US-0134219
PR   14-MAY-1999;   99US-0134221
PR   14-MAY-1999;   99US-0134370
PR   18-MAY-1999;   99US-0134768
PR   19-MAY-1999;   99US-0134941
PR   20-MAY-1999;   99US-0135124
PR   21-MAY-1999;   99US-0135353
PR   24-MAY-1999;   99US-0135629
PR   25-MAY-1999;   99US-0136021
PR   27-MAY-1999;   99US-0136392
PR   28-MAY-1999;   99US-0136782
PR   01-JUN-1999;   99US-0137222
PR   03-JUN-1999;   99US-0137528
PR   04-JUN-1999;   99US-0137502
PR   07-JUN-1999;   99US-0138724
PR   08-JUN-1999;   99US-0139094
PR   10-JUN-1999;   99US-0139540
PR   10-JUN-1999;   99US-0139647
PR   14-JUN-1999;   99US-0140119
PR   16-JUN-1999;   99US-0139452
PR   16-JUN-1999;   99US-0139453
PR   17-JUN-1999;   99US-0139492
PR   18-JUN-1999;   99US-0139454
PR   18-JUN-1999;   99US-0139455
PR   18-JUN-1999;   99US-0139456
PR   18-JUN-1999;   99US-0139457
PR   18-JUN-1999;   99US-0139458
PR   18-JUN-1999;   99US-0139459
PR   18-JUN-1999;   99US-0139460
PR   18-JUN-1999;   99US-0139461
PR   18-JUN-1999;   99US-0139462
PR   18-JUN-1999;   99US-0139463
PR   18-JUN-1999;   99US-0139750
PR   18-JUN-1999;   99US-0139763
PR   21-JUN-1999;   99US-0139817
PR   22-JUN-1999;   99US-0139899
PR   23-JUN-1999;   99US-0140353
PR   23-JUN-1999;   99US-0140354
PR   24-JUN-1999;   99US-0140695
PR   28-JUN-1999;   99US-0140823
PR   29-JUN-1999;   99US-0140991
PR   30-JUN-1999;   99US-0141287
PR   01-JUL-1999;   99US-0141842
PR   01-JUL-1999;   99US-0142154
PR   02-JUL-1999;   99US-0142055
PR   06-JUL-1999;   99US-0142390
```

PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 4.4%; Score 43; DB 21; Length 289;
Best Local Similarity 100.0%; Pred. No. 1.3e-35;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      805 VLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQGFLKMLAG 847
            |||||||||||||||||||||||||||||||||||||||||||
Db      123 VLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQGFLKMLAG 165
```

RESULT 11
AAG50461
ID AAG50461 standard; Protein; 296 AA.
XX
AC AAG50461;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 63952.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.

R	25-MAR-1999;	99US-0126264.
R	29-MAR-1999;	99US-0126785.
R	01-APR-1999;	99US-0127462.
R	06-APR-1999;	99US-0128234.
R	08-APR-1999;	99US-0128714.
R	16-APR-1999;	99US-0129845.
R	19-APR-1999;	99US-0130077.
R	21-APR-1999;	99US-0130449.
R	23-APR-1999;	99US-0130510.
R	23-APR-1999;	99US-0130891.
R	28-APR-1999;	99US-0131449.
R	30-APR-1999;	99US-0132048.
R	30-APR-1999;	99US-0132407.
R	04-MAY-1999;	99US-0132484.
R	05-MAY-1999;	99US-0132485.
R	06-MAY-1999;	99US-0132486.
R	06-MAY-1999;	99US-0132487.
R	07-MAY-1999;	99US-0132863.
R	11-MAY-1999;	99US-0134256.
R	14-MAY-1999;	99US-0134218.
R	14-MAY-1999;	99US-0134219.
R	14-MAY-1999;	99US-0134221.
R	14-MAY-1999;	99US-0134370.
R	18-MAY-1999;	99US-0134768.
R	19-MAY-1999;	99US-0134941.
R	20-MAY-1999;	99US-0135124.
R	21-MAY-1999;	99US-0135353.
R	24-MAY-1999;	99US-0135629.
R	25-MAY-1999;	99US-0136021.
R	27-MAY-1999;	99US-0136392.
R	28-MAY-1999;	99US-0136782.
R	01-JUN-1999;	99US-0137222.
R	03-JUN-1999;	99US-0137528.
R	04-JUN-1999;	99US-0137502.
R	07-JUN-1999;	99US-0137724.
R	08-JUN-1999;	99US-0138094.
R	10-JUN-1999;	99US-0138540.
R	10-JUN-1999;	99US-0138847.
R	14-JUN-1999;	99US-0139119.
R	16-JUN-1999;	99US-0139452.
R	16-JUN-1999;	99US-0139453.
R	17-JUN-1999;	99US-0139492.
R	18-JUN-1999;	99US-0139454.
R	18-JUN-1999;	99US-0139455.
R	18-JUN-1999;	99US-0139456.
R	18-JUN-1999;	99US-0139457.
R	18-JUN-1999;	99US-0139458.
R	18-JUN-1999;	99US-0139459.
R	18-JUN-1999;	99US-0139460.
R	18-JUN-1999;	99US-0139461.
R	18-JUN-1999;	99US-0139462.
R	18-JUN-1999;	99US-0139463.
R	18-JUN-1999;	99US-0139750.
R	18-JUN-1999;	99US-0139763.
R	21-JUN-1999;	99US-0139817.
R	22-JUN-1999;	99US-0139899.
R	23-JUN-1999;	99US-0140353.
R	23-JUN-1999;	99US-0140354.
R	24-JUN-1999;	99US-0140695.
R	28-JUN-1999;	99US-0140823.
PR	29-JUN-1999;	99US-0140991.
PR	30-JUN-1999;	99US-0141287.
PR	01-JUL-1999;	99US-0141842.
PR	01-JUL-1999;	99US-0142154.
PR	02-JUL-1999;	99US-0142055.
PR	06-JUL-1999;	99US-0142390.
PR	08-JUL-1999;	99US-0142803.
PR	09-JUL-1999;	99US-0142920.
PR	12-JUL-1999;	99US-0142977.
PR	13-JUL-1999;	99US-0143542.
PR	14-JUL-1999;	99US-0143624.
PR	15-JUL-1999;	99US-0144005.
PR	16-JUL-1999;	99US-0144085.
PR	16-JUL-1999;	99US-0144086.
PR	19-JUL-1999;	99US-0144325.
PR	19-JUL-1999;	99US-0144331.
PR	19-JUL-1999;	99US-0144332.
PR	19-JUL-1999;	99US-0144333.
PR	19-JUL-1999;	99US-0144334.
PR	19-JUL-1999;	99US-0144335.
PR	20-JUL-1999;	99US-0144352.
PR	20-JUL-1999;	99US-0144632.
PR	20-JUL-1999;	99US-0144884.
PR	21-JUL-1999;	99US-0145814.
PR	21-JUL-1999;	99US-0146086.
PR	21-JUL-1999;	99US-0146088.
PR	22-JUL-1999;	99US-0146085.
PR	22-JUL-1999;	99US-0146087.
PR	22-JUL-1999;	99US-0146089.
PR	22-JUL-1999;	99US-0146197.
PR	23-JUL-1999;	99US-0146145.
PR	23-JUL-1999;	99US-0146218.
PR	23-JUL-1999;	99US-0146224.
PR	26-JUL-1999;	99US-0146230.
PR	27-JUL-1999;	99US-0146315.
PR	27-JUL-1999;	99US-0146316.
PR	27-JUL-1999;	99US-0146319.
PR	28-JUL-1999;	99US-0146951.
PR	02-AUG-1999;	99US-0146386.
PR	02-AUG-1999;	99US-0146388.
PR	02-AUG-1999;	99US-0146389.
PR	03-AUG-1999;	99US-0147036.
PR	04-AUG-1999;	99US-0147204.
PR	04-AUG-1999;	99US-0147302.
PR	05-AUG-1999;	99US-0147192.
PR	05-AUG-1999;	99US-0147260.
PR	06-AUG-1999;	99US-0147303.
PR	06-AUG-1999;	99US-0147416.
PR	09-AUG-1999;	99US-0147493.
PR	09-AUG-1999;	99US-0147935.
PR	10-AUG-1999;	99US-0148171.
PR	11-AUG-1999;	99US-0148319.
PR	12-AUG-1999;	99US-0148341.
PR	13-AUG-1999;	99US-0148565.
PR	13-AUG-1999;	99US-0148684.
PR	16-AUG-1999;	99US-0149368.
PR	17-AUG-1999;	99US-0149175.
PR	18-AUG-1999;	99US-0149426.
PR	20-AUG-1999;	99US-0149722.
PR	20-AUG-1999;	99US-0149723.
PR	20-AUG-1999;	99US-0149929.
PR	23-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0153363.
PR	10-SEP-1999;	99US-0153970.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.

PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 4.4%; Score 43; DB 21; Length 296;
Best Local Similarity 100.0%; Pred. No. 1.3e-35;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 805 VLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQGFLKMLSG 847
||||||||||||||||||||||||||||||||||||||||||||
Db 122 VLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQGFLKMLSG 164

RESULT 12
AAB16338
ID AAB16338 standard; Protein; 955 AA.
XX
AC AAB16338;
XX
DT 31-OCT-2000 (first entry)
XX
DE Eucalyptus grandis cellulose synthase protein sequence SEQ ID NO:145.
XX
KW Eucalyptus grandis; pinus radiata; Monterey pine; modification;
KW plant cell wall; polysaccharide; polysaccharide biosynthetic pathway;
KW transgenic plant.
XX
OS Eucalyptus grandis.
XX
PN WO200022092-A2.
XX
PD 20-APR-2000.
XX
PF 08-OCT-1999; 99WO-NZ00169.
XX
PR 13-OCT-1998; 98US-0170862.
PR 11-AUG-1999; 99US-0148426.
XX
PA (GENE-) GENESIS RES & DEV CORP LTD.
PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI Bloksberg LN;
XX
DR WPI; 2000-339328/29.
DR N-PSDB; AAA67145.
XX
PT New genes encoding proteins involved in a plant polysaccharide
PT biosynthetic pathway, useful for modulating or altering the
PT polysaccharide content, composition or structure of the plant -
XX
PS Claim 17; Page 111-113; 301pp; English.
XX
CC The present invention describes isolated polynucleotides (PN) comprising
CC a sequence selected from one of 835 nucleotide sequences given in
CC AAA67073 to AAA67907, their (reverse) complements, sequences producing
CC an Expectation (E) value of 0.01 or less compared to the 835 sequences,
CC sequences at least 50% identical to them, 200, 100, 40 or 20-mers of the
CC 835 sequences or sequences that are degenerately equivalent or allelic
CC to the 835 sequences. The polynucleotides are used to modify the
CC activity of a polypeptide involved in a polysaccharide biosynthetic
CC pathway in the plant. They are especially used to modulate or alter the
CC polysaccharide content, composition or structure of the plant. AAB16268
CC to AAB16340 are proteins encoded by some of the polynucleotide sequence
CC given in the present invention.
XX
SQ Sequence 955 AA;

Query Match 4.4%; Score 43; DB 21; Length 955;
Best Local Similarity 100.0%; Pred. No. 3.8e-35;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 891 AGFSDALNKGYEAWGPLFGKVFFAFWVILHLYPFLKGLMGRQN 933
|||||||||||||||||||||||||||||||||||||||||||||
Db 892 AGFSDALNKGYEAWGPLFGKVFFAFWVILHLYPFLKGLMGRQN 934

RESULT 13
AAB16279
ID AAB16279 standard; Protein; 239 AA.
XX
AC AAB16279;
XX
DT 31-OCT-2000 (first entry)
XX
DE Eucalyptus grandis cellulose synthase protein sequence SEQ ID NO:41.
XX
KW Eucalyptus grandis; pinus radiata; Monterey pine; modification;
KW plant cell wall; polysaccharide; polysaccharide biosynthetic pathway;
KW transgenic plant.
XX
OS Eucalyptus grandis.
XX
PN WO200022092-A2.
XX
PD 20-APR-2000.
XX
PF 08-OCT-1999; 99WO-NZ00169.
XX
PR 13-OCT-1998; 98US-0170862.
PR 11-AUG-1999; 99US-0148426.
XX
PA (GENE-) GENESIS RES & DEV CORP LTD.
PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI Bloksberg LN;
XX
DR WPI; 2000-339328/29.
DR N-PSDB; AAA67084.
XX
PT New genes encoding proteins involved in a plant polysaccharide
PT biosynthetic pathway, useful for modulating or altering the
PT polysaccharide content, composition or structure of the plant -
XX
PS Claim 17; Page 55-56; 301pp; English.
XX
CC The present invention describes isolated polynucleotides (PN) comprising
CC a sequence selected from one of 835 nucleotide sequences given in
CC AAA67073 to AAA67907, their (reverse) complements, sequences producing
CC an Expectation (E) value of 0.01 or less compared to the 835 sequences,

sequences at least 50% identical to them, 200, 100, 40 or 20-mers of the 835 sequences or sequences that are degenerately equivalent or allelic to the 835 sequences. The polynucleotides are used to modify the activity of a polypeptide involved in a polysaccharide biosynthetic pathway in the plant. They are especially used to modulate or alter the polysaccharide content, composition or structure of the plant. AAB16268 to AAB16340 are proteins encoded by some of the polynucleotide sequence given in the present invention.

) Sequence 239 AA;

Query Match 4.1%; Score 40; DB 21; Length 239;
Best Local Similarity 100.0%; Pred. No. 1.4e-32;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 256 FVSTVDPLKEPPLITANTVLSILAVDYPVDKVSCYVSDDG 295
||||||||||||||||||||||||||||||||||||||||
D 20 FVSTVDPLKEPPLITANTVLSILAVDYPVDKVSCYVSDDG 59

ESULT 14
AW46202
D AAW46202 standard; Protein; 547 AA.
X
C AAW46202;
X
T 06-JUL-1998 (first entry)
X
E Rice RSW1-like cellulose synthase.
X
K Cellulose synthase; RSW1-like protein; beta-1,4-glucan;
W transgenic plant; rice.
X
S Oryza sativa.
X
N WO9800549-A1.
X
D 08-JAN-1998.
X
F 24-JUN-1997; 97WO-AU00402.
X
R 27-JUN-1996; 96AU-0000699.
X
A (CSIR) COMMONWEALTH SCI & IND RES ORG.
A (AUSU) UNIV AUSTRALIAN NAT.
X
I Arioli A, Betzner AS, Peng L, Williamson RE;
X
R WPI; 1998-086974/08.
R N-PSDB; AAV06569.
X
T DNA encoding cellulose biosynthetic enzyme - useful for manipulation
T of cellulose and beta-1,4-glucan
X
S Claim 29; Page 167-169; 207pp; English.
X
C This polypeptide comprises the N-terminal portion of an RSW1-like
C cellulose synthase of rice. It shows 70% identity to Arabidopsis
C thaliana RSW1 cellulose synthase (see AAW33817). The amino acid
C sequence was deduced from partial cDNA clone EST S0542 that was
C isolated on the basis of homology to Arabidopsis RSW1 nuclectide
'C sequences (see AAV06562-68). Claimed nucleic acid molecules (see
C AAV06562-69) coding for claimed polypeptides (see AAW33816-20 and
'C AAW46202) involved in cellulose biosynthesis can be used to
'C manipulate the cellulose and/or beta-glucan content of transgenic
'C plants. Expression of nucleic acids in the sense orientation
'C increases the level of cellulose and reduces the level of
'C non-crystalline beta-1,4-glucan and starch, providing plants with
'C modified strength and/or shape and/or fibre properties, or having
'C increased resistance to stresses or pests. Antisense, ribozyme or
CC co-suppression molecules can be used to reduce the cellulose
CC content of a transgenic plant, e.g. to improve digestibility or to
CC alter carbon partitioning such that increased carbon is available
CC for growth.
XX
SQ Sequence 547 AA;

Query Match 4.1%; Score 40; DB 19; Length 547;
Best Local Similarity 100.0%; Pred. No. 3.1e-32;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 256 FVSTVDPLKEPPLITANTVLSILAVDYPVDKVSCYVSDDG 295
||||||||||||||||||||||||||||||||||||||||
Db 354 FVSTVDPLKEPPLITANTVLSILAVDYPVDKVSCYVSDDG 393

RESULT 15
AAY84110
ID AAY84110 standard; Protein; 1074 AA
XX
AC AAY84110;
XX
DT 03-JUL-2000 (first entry)
XX
DE Amino acid sequence of [illegible] cellulose synthase.
XX
KW Maize; cellulose synthase; stalk quality; stand; silage; cellulose;
KW transgenic plant; plant breeding marker.
XX
OS Zea mays.
XX
PN WO200009706-A2
XX
PD 24-FEB-2000.
XX
PF 16-AUG-1999; 99WO-US18769
XX
PR 17-AUG-1998; 98US-0096822.
XX
PA (PION-) PIONEER HI BRED INT INC
XX
PI Dhugga KS, Helentjaris TG, Bowen BA, Wang X;
XX
DR WPI; 2000-224343/19.
DR N-PSDB; AAZ99500.
XX
PT New genes which encode maize cellulose synthase polypeptides in plants
PT useful for modulating the expression of cellulose synthase in plants
PT and to produce transgenic plants expressing the novel protein -
XX
PS Claim 15; Page 110-113; 119pp; English
XX
CC The present sequence represents a maize cellulose synthase polypeptide.
CC The cellulose synthase can be used for the improvement of stalk quality
CC for improved stand or silage. It also provides an increased concentration
CC of cellulose in the pericarp, hardening the kernel and improving its
CC handling ability. The sequences are used to produce transgenic plants
CC and seeds expressing the cellulose synthase. The polynucleotide is
CC used for modulating, preferably increasing, the level of the synthase
CC in a plant cell. The plants are preferably monocots. The polynucleotide
CC is also used as a probe or primer in the detection quantitation or
CC isolation of gene transcripts. The probes are useful in detecting
CC deficiencies in the level of mRNA in screenings for desired transgenic
CC plant, for detecting mutations in the gene, for monitoring upregulation
CC of expression or change in enzyme activity in screening assays of
CC compounds, for detection of any number of allelic variants of the gene,
CC or for use as molecular markers in plant breeding programs. The
CC isolated nucleic acids of the present invention can also be used for
CC recombinant expression of their encoded polypeptides or for use as
CC immunogens in the preparation and/or screening of antibodies. The
CC proteins can be employed in assays for enzyme agonists or antagonists
CC of enzyme function or for use of immunogens or antigens to obtain
CC antibodies specifically immunoreactive with a protein.
XX
SQ Sequence 1074 AA;

Query Match 4.1%; Score 40; DB 21; Length 1074;
Best Local Similarity 100.0%; Pred. No. 5.6e-32;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 256 PVSTVDPLKEPPLITANTVLSILAVDYPVDKVSCYVSDDG 295
Db 352 PVSTVDPLKEPPLITANTVLSILAVDYPVDKVSCYVSDDG 391

Search completed: October 29, 2003, 18:34:27
Job time : 76.6673 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 31, 2003, 09:56:20 ; Search time 155.118 Seconds
(without alignments)
2782.667 Million cell updates/sec

Title: US-09-980-043-2
Perfect score: 5240
Sequence: 1 MMESGAPICHTCGEQVGHDA.....PFVNKVDNTLVAETCISIDC 978

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09980043/runat_29102003_163853_21883/app_query.fasta_1.1358
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human4C.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09980043_@CGN_1_1_76_@runat_29102003_163853_21883 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4540	86.6	3328	3	US-08-960-048-1	Sequence 1, Appli
2	4540	86.6	3328	4	US-09-838-586-1	Sequence 1, Appli
3	3350	63.9	3603	4	US-09-221-013A-5	Sequence 5, Appli
4	3346	63.9	3673	4	US-09-221-013A-11	Sequence 11, Appl
5	3329	63.5	3614	4	US-09-221-013A-9	Sequence 9, Appli
6	3242	61.9	3828	4	US-09-221-013A-7	Sequence 7, Appli
7	2925.5	55.8	8411	4	US-09-221-013A-3	Sequence 3, Appli
8	2311	44.1	5009	4	US-09-221-013A-4	Sequence 4, Appli
9	2284.5	43.6	2248	4	US-09-221-013A-1	Sequence 1, Appli
10	1760	33.6	4612	3	US-08-960-048-2	Sequence 2, Appli
11	1760	33.6	4612	4	US-09-838-586-2	Sequence 2, Appli
12	1396.5	26.7	1741	4	US-09-221-013A-13	Sequence 13, Appl
13	383.5	7.3	291	4	US-09-313-294A-5519	Sequence 5519, Ap
14	377	7.2	281	4	US-09-313-294A-6485	Sequence 6485, Ap
15	363.5	6.9	300	4	US-09-313-294A-4753	Sequence 4753, Ap
16	363	6.9	268	4	US-09-313-294A-849	Sequence 849, App
17	325.5	6.2	283	4	US-09-313-294A-1981	Sequence 1981, Ap
18	319	6.1	286	4	US-09-313-294A-4696	Sequence 4696, Ap
19	287.5	5.5	287	4	US-09-313-294A-5274	Sequence 5274, Ap
20	268	5.1	284	4	US-09-313-294A-3993	Sequence 3993, Ap
21	252	4.8	278	4	US-09-313-294A-4015	Sequence 4015, Ap
22	223	4.3	2262	1	US-07-689-008-3	Sequence 3, Appli
23	223	4.3	9540	1	US-07-689-008-1	Sequence 1, Appli
24	217	4.1	16836	4	US-09-147-236-1	Sequence 1, Appli
25	217	4.1	16836	4	US-09-147-236-10	Sequence 10, Appl
26	217	4.1	16836	4	US-09-522-474-1	Sequence 1, Appli
27	217	4.1	16836	4	US-09-522-474-10	Sequence 10, Appl
28	204.5	3.9	250	4	US-09-313-294A-582	Sequence 582, App
29	198.5	3.8	159	4	US-09-313-294A-1693	Sequence 1693, Ap
30	197	3.8	317	4	US-09-313-294A-7123	Sequence 7123, Ap
31	174.5	3.3	2912	4	US-08-541-939-1	Sequence 1, Appli
32	174.5	3.3	2912	5	PCT-US91-01726-2	Sequence 2, Appli
33	133	2.5	3534	4	US-09-795-927-1	Sequence 1, Appli
34	133	2.5	6080	4	US-09-620-312D-327	Sequence 327, App
35	133	2.5	7277	4	US-09-795-927-5	Sequence 5, Appli
36	121.5	2.3	591	4	US-09-276-599-7	Sequence 7, Appli
37	116	2.2	480	4	US-09-276-599-8	Sequence 8, Appli
38	115.5	2.2	5822	4	US-09-354-147C-4	Sequence 4, Appli
39	114.5	2.2	2617	3	US-09-161-443-1	Sequence 1, Appli
40	114	2.2	3407	1	US-08-253-155A-7	Sequence 7, Appli
41	113.5	2.2	1506	4	US-09-328-352-1101	Sequence 1101, Ap
42	113.5	2.2	11601	2	US-08-222-617A-3	Sequence 3, Appli
43	113.5	2.2	11601	2	US-08-222-617A-24	Sequence 24, Appl
44	112	2.1	6412	3	US-08-652-877-17	Sequence 17, Appl
45	112	2.1	6412	3	US-08-476-515A-17	Sequence 17, Appl

ALIGNMENTS

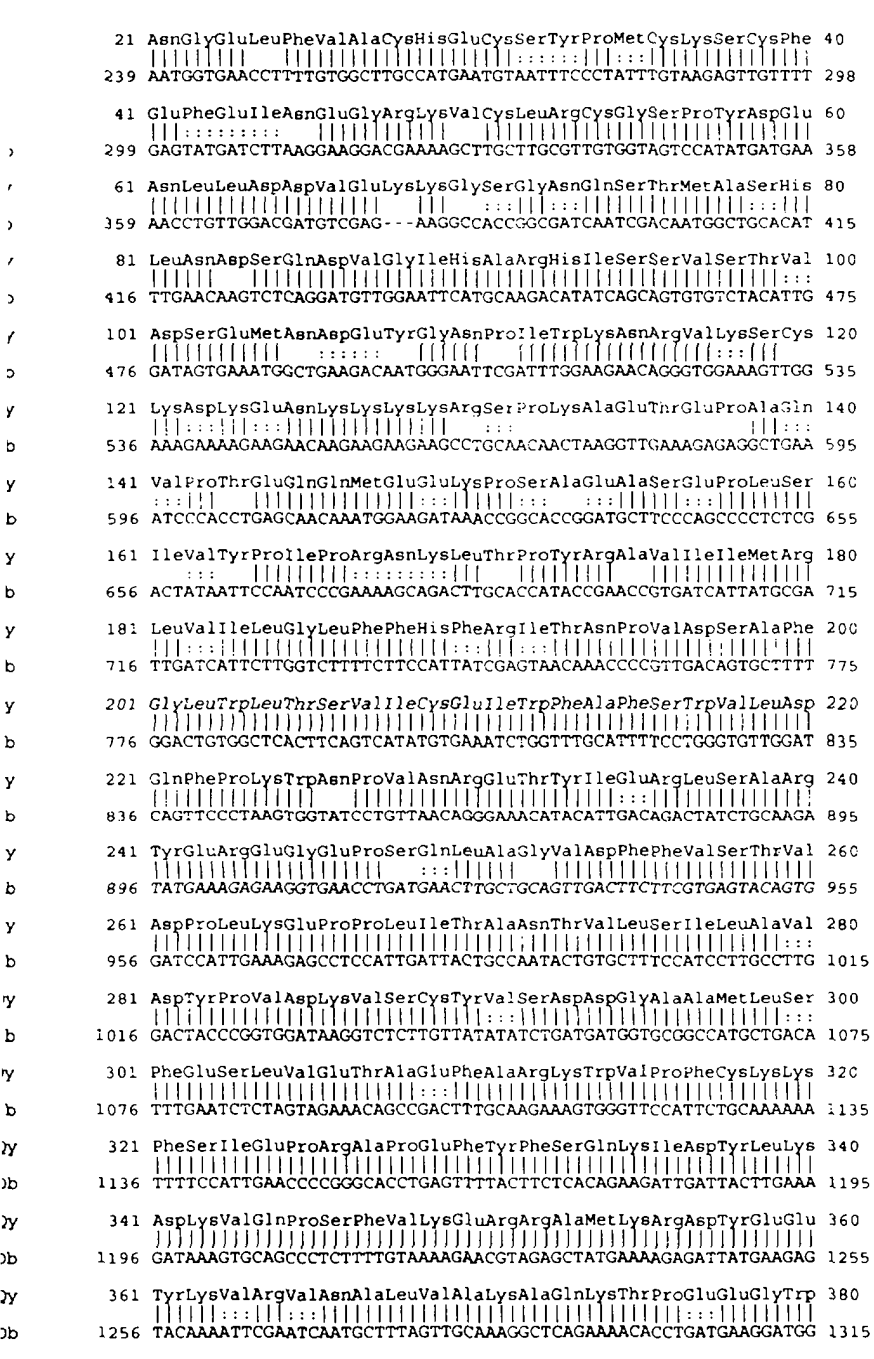
RESULT 1
US-08-960-048-1
; Sequence 1, Application US/08960048C
; Patent No. 6271443
; GENERAL INFORMATION:
; APPLICANT: Stalker, D. et al.
; TITLE OF INVENTION: Plant Cellulose Synthase and Promoter
; TITLE OF INVENTION: Sequences
; FILE REFERENCE: 15621/01/US
; CURRENT APPLICATION NUMBER: US/08/960,048C
; CURRENT FILING DATE: 1997-10-29
; PRIOR APPLICATION NUMBER: 60/029,987
; PRIOR FILING DATE: 1996-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 3328
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Oligonucleotide
US-08-960-048-1

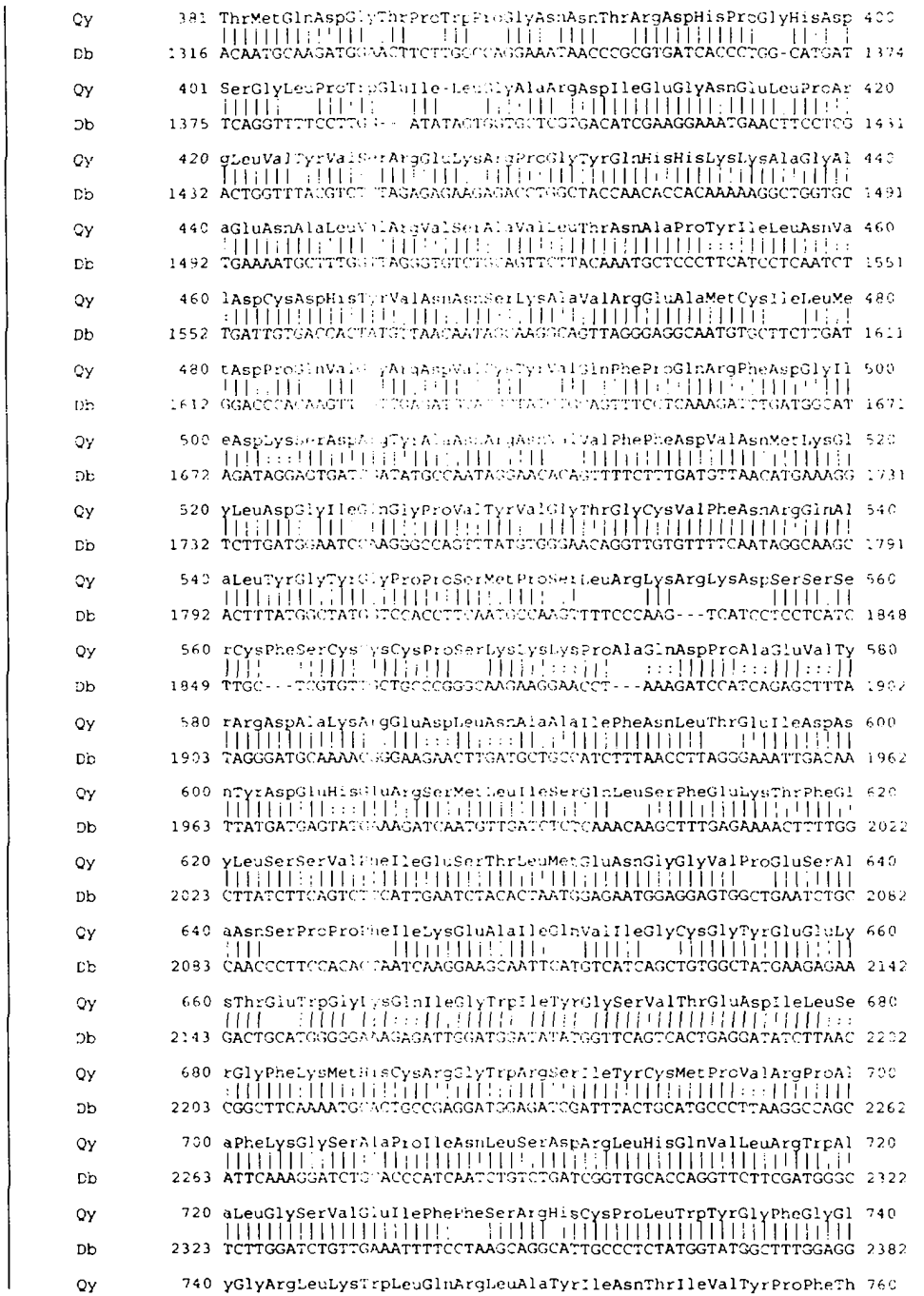
Alignment Scores:

Pred. No.:	0	Length:	3328
Score:	4540.00	Matches:	845
Percent Similarity:	93.27%	Conservative:	69
Best Local Similarity:	86.22%	Mismatches:	59
Query Match:	86.64%	Indels:	8
DB:	3	Gaps:	6

US-09-980-043-2 (1-978) x US-08-960-048-1 (1-3328)

Qy 1 MetMetGluSerGlyAlaProIleCysHisThrCysGlyGluGlnValGlyHisAspAla 20
Db 179 ATGATGGAATCTGGGGTTCCTGTTTGCCACACTTGTGGTGAACATGTTGGGTTGAATGTT 238


21 AsnGlyGluLeuPheValAlaCysHisGluCysSerTyrProMetCysLysSerCysPhe 40
239 AATGGTGAACCTTTTGTGGCTTGCCATGAATGTAATTTCCCTATTTGTAAGAGTTGTTTT 298
41 GluPheGluIleAsnGluGlyArgLysValCysLeuArgCysGlySerProTyrAspGlu 60
299 GAGTATGATCTTAAGGAAGGACGAAAAGCTTGCTTGCGTTGTGGTAGTCCATATGATGAA 358
61 AsnLeuLeuAspAspValGluLysLysGlySerGlyAsnGlnSerThrMetAlaSerHis 80
359 AACCTGTTGGACGATGTCGAG---AAGGCCACCGGCGATCAATCGACAATGGCTGCACAT 415
81 LeuAsnAspSerGlnAspValGlyIleHisAlaArgHisIleSerSerValSerThrVal 100
416 TTGAACAAGTCTCAGGATGTTGGAATTCATGCAAGACATATCAGCAGTGTGTCTACATTG 475
101 AspSerGluMetAsnAspGluTyrGlyAsnProIleTrpLysAsnArgValLysSerCys 120
476 GATAGTGAAATGGCTGAAGACAATGGGAATTCGATTTGGAAGAACAGGGTGGAAAGTTGG 535
121 LysAspLysGluAsnLysLysLysLysArgSerProLysAlaGluThrGluProAlaGln 140
536 AAAGAAAAGAAGAACAAGAAGAAGAAGCCTGCAACAACTAAGGTTGAAAGAGAGGCTGAA 595
141 ValProThrGluGlnGlnMetGluGluLysProSerAlaGluAlaSerGluProLeuSer 160
596 ATCCCACCTGAGCAACAAATGGAAGATAAACCGGCACCGGATGCTTCCCAGCCCCTCTCG 655
161 IleValTyrProIleProArgAsnLysLeuThrProTyrArgAlaValIleIleMetArg 180
656 ACTATAATTCCAATCCCGAAAAGCAGACTTGCACCATACCGAACCGTGATCATTATGCGA 715
181 LeuValIleLeuGlyLeuPhePheHisPheArgIleThrAsnProValAspSerAlaPhe 200
716 TTGATCATTCTTGGTCTTTTCTTCCATTATCGAGTAACAAACCCCGTTGACAGTGCTTTT 775
201 GlyLeuTrpLeuThrSerValIleCysGluIleTrpPheAlaPheSerTrpValLeuAsp 220
776 GGACTGTGGCTCACTTCAGTCATATGTGAAATCTGGTTTGCATTTTCCTGGGTGTTGGAT 835
221 GlnPheProLysTrpAsnProValAsnArgGluThrTyrIleGluArgLeuSerAlaArg 240
836 CAGTTCCCTAAGTGGTATCCTGTTAACAGGGAAACATACATTGACAGACTATCTGCAAGA 895
241 TyrGluArgGluGlyGluProSerGlnLeuAlaGlyValAspPhePheValSerThrVal 260
896 TATGAAAGAGAAGGTGAACCTGATGAACTTGCTGCAGTTGACTTCTTCGTGAGTACAGTG 955
261 AspProLeuLysGluProProLeuIleThrAlaAsnThrValLeuSerIleLeuAlaVal 280
956 GATCCATTGAAAGAGCCTCCATTGATTACTGCCAATACTGTGCTTTCCATCCTTGCCTTG 1015
281 AspTyrProValAspLysValSerCysTyrValSerAspAspGlyAlaAlaMetLeuSer 300
1016 GACTACCCGGTGGATAAGGTCTCTTGTTATATATCTGATGATGGTGCGGCCATGCTGACA 1075
301 PheGluSerLeuValGluThrAlaGluPheAlaArgLysTrpValProPheCysLysLys 320
1076 TTTGAATCTCTAGTAGAAACAGCCGACTTTGCAAGAAAGTGGGTTCCATTCTGCAAAAAA 1135
321 PheSerIleGluProArgAlaProGluPheTyrPheSerGlnLysIleAspTyrLeuLys 340
1136 TTTTCCATTGAACCCCGGGCACCTGAGTTTTACTTCTCACAGAAGATTGATTACTTGAAA 1195
341 AspLysValGlnProSerPheValLysGluArgArgAlaMetLysArgAspTyrGluGlu 360
1196 GATAAAGTGCAGCCCTCTTTTGTAAAAGAACGTAGAGCTATGAAAAGAGATTATGAAGAG 1255
361 TyrLysValArgValAsnAlaLeuValAlaLysAlaGlnLysThrProGluGluGlyTrp 380
1256 TACAAAATTCGAATCAATGCTTTAGTTGCAAAGGCTCAGAAAACACCTGATGAAGGATGG 1315


Qy 381 ThrMetGlnAspGlyThrProTrpProGlyAsnAsnThrArgAspHisProGlyHisAsp 400
Db 1316 ACAATGCAAGATGGTACTTCTTGCCCAGGAAATAACCCGCGTGATCACCCTGG-CATGAT 1374
Qy 401 SerGlyLeuProTrpGluIle-LeuGlyAlaArgAspIleGluGlyAsnGluLeuProAr 420
Db 1375 TCAGGTTTCCTTGG-- ATATAGTGGTGCTCGTGACATCGAAGGAAATGAACTTCCTCG 1431
Qy 421 gLeuValTyrValSerArgGluLysArgProGlyTyrGlnHisHisLysLysAlaGlyAl 440
Db 1432 ACTGGTTTACGTCT TAGAGAGAAGAGACCTGGCTACCAACACCACAAAAAGGCTGGTGC 1491
Qy 441 aGluAsnAlaLeuVa lArgValSerAlaValLeuThrAsnAlaProTyrIleLeuAsnVa 460
Db 1492 TGAAAATGCTTTGG TAGGGTGTCTTCAGTTCTTACAAATGCTCCCTTCATCCTCAATCT 1551
Qy 461 lAspCysAspHisTyrValAsnAsnSerLysAlaValArgGluAlaMetCysIleLeuMe 480
Db 1552 TGATTGTGACCACTATGTTAACAATAGCAAGGCAGTTAGGGAGGCAATGTGCTTCTTGAT 1611
Qy 481 tAspProGlnVal yArgAspVal TyrTyrValGlnPheProGlnArgPheAspGlyIl 500
Db 1612 GGACCCACAAGTT TGAGATTA TAGCAGTTTCCTCAAAGATTTGATGGCAT 1671
Qy 501 eAspLysSerAspA gTyrAlaAsnArgAsnValPhePheAspValAsnMetLysGl 520
Db 1672 AGATAGGAGTGAT ATATGCCAATAGGAACACAGTTTCTTTGATGTTAACATGAAAGG 1731
Qy 521 yLeuAspGlyIleGlnGlyProValTyrValGlyThrGlyCysValPheAsnArgGlnAl 540
Db 1732 TCTTGATGGAATCCAAGGGCCAGTTTATGTGGGAACAGGTTGTGTTTTCAATAGGCAAGC 1791
Qy 541 aLeuTyrGlyTyrGlyProProSerMetProMetLeuArgLysArgLysAspSerSerSe 560
Db 1792 ACTTTATGGCTATG TCCACCTTCAATGCCAAGTTTTCCCAAG---TCATCCTCCTCATC 1848
Qy 561 rCysPheSerCys ysCysProSerLysLysLysProAlaGlnAspProAlaGluValTy 580
Db 1849 TTGC---TGGTGTT CTGCCCGGGCAAGAAGGAACCT---AAAGATCCATCAGAGCTTTA 1902
Qy 581 rArgAspAlaLysArgGluAspLeuAsnAlaAlaIlePheAsnLeuThrGluIleAspAs 600
Db 1903 TAGGGATGCAAAAC GGAAGAACTTGATGCTGCCATCTTTAACCTTAGGGAAATTGACAA 1962
Qy 601 nTyrAspGluHisGluArgSerMetLeuIleSerGlnLeuSerPheGluLysThrPheGl 620
Db 1963 TTATGATGAGTATG AAGATCAATGTTGATCTCTCAAACAAGCTTTGAGAAAACTTTTGG 2022
Qy 621 yLeuSerSerValPheIleGluSerThrLeuMetGluAsnGlyGlyValProGluSerAl 640
Db 2023 CTTATCTTCAGTCT CATTGAATCTACACTAATGGAGAATGGAGGAGTGGCTGAATCTGC 2082
Qy 641 aAsnSerProProPheIleLysGluAlaIleGlnValIleGlyCysGlyTyrGluGluLy 660
Db 2083 CAACCCTTCCACAC AATCAAGGAAGCAATTCATGTCATCAGCTGTGGCTATGAAGAGAA 2142
Qy 661 sThrGluTrpGlyLysGlnIleGlyTrpIleTyrGlySerValThrGluAspIleLeuSe 680
Db 2143 GACTGCATGGGGGAAAGAGATTGGATGGATATATGGTTCAGTCACTGAGGATATCTTAAC 2202
Qy 681 rGlyPheLysMetHisCysArgGlyTrpArgSerIleTyrCysMetProValArgProAl 700
Db 2203 CGGCTTCAAAATG ACTGCCGAGGATGGAGATCGATTTACTGCATGCCCTTAAGGCCAGC 2262
Qy 701 aPheLysGlySerAlaProIleAsnLeuSerAspArgLeuHisGlnValLeuArgTrpAl 720
Db 2263 ATTCAAAGGATCTG ACCCATCAATCTGTCTGATCGGTTGCACCAGGTTCTTCGATGGGC 2322
Qy 721 aLeuGlySerValGluIlePhePheSerArgHisCysProLeuTrpTyrGlyPheGlyGl 740
Db 2323 TCTTGGATCTGTTGAAATTTTCCTAAGCAGGCATTGCCCTCTATGGTATGGCTTTGGAGG 2382
Qy 741 yGlyArgLeuLysTrpLeuGlnArgLeuAlaTyrIleAsnThrIleValTyrProPheTh 760

```
Db       2383 TGGTCGTCTTAAATGGCTTCAAAGACTAGCATATATAAACA CATTGTCTATCCTTTCAC 2442

Qy        760 rSerLeuProLeuIleAlaTyrCysThrIleProAlaValCysLeuLeuThrGlyLysPh 780

Db       2443 ATCCCTTCCACTCATTGCCTATTGTTCACTACCAGCAATCT TCTTCTCACAGGAAAATT 2502

Qy        780 eIleIleProThrLeuSerAsnLeuAlaSe:MetLeuPheLeuGlyLeuPheIleSerIl 800

Db       2503 TATCATACCAACGCTCTCAAACC GGCAAGTGTTCTCTTTC TGGCCTTTTCCTTTCCAT 2562

Qy        800 eIleValThrAlaValLeuGluLeuArgTrpSerGlyVal Ser IleGluAspLeuTrpAr 820

Db       2563 TATCGTGACTGCTGTTCTCGAGCTCCGATGGAGTGGTGTCA CATTGAGGACTTATGGGG 2622

Qy        820 gAsnGluGlnPheTrpValIleGlyGlyValSerAlaHisLeuPheAlaValPheGlnGl 840

Db       2623 TAACGAGCAGTTTTGGGTCATCGGTGGCGTTTCAGCCCATCTCTTTGCCGTCTTCCAAGG 2682

Qy        840 yPheLeuLysMetLeuAlaGlyIleAspThrAsnPheThrValThrAlaLysAlaAlaGl 860

Db       2683 TTTCCTTAAGATGCTTGCGGGCATTGACACCAACTTTACTGTCACTGCCAAAGCAGCTGA 2742

Qy        860 uAspAlaGluPheGlyGluLeuTyrMetValLysTrpThrThrLeuLeuIleProProTh 880

Db       2743 TGATGCAGATTTTGGTGAGCTCTACATTGTGAAATGGACTA ACTTCTAATCCCTCCAAC 2802

Qy        880 rThrLeuLeuIleIleAsnMetSerGly---CysAlaGlyPheSerAspAlaLeuAsnLy 899

Db       2803 AACACTCCTCATCGTCAACATGGTTGGTGTCGTTGCCGGATTCTCCGATGCCCTCAACAA 2862

Qy        899 sGlyTyrGluAlaTrpGlyProLeuPheGlyLysValPhePheAlaPheTrpValIleLe 919

Db       2863 AGGGTACGAAGCTTGGGGACCACT GTTTGGCAAAGTGTTCT TTCCTTCTGGGTCATCCT 2922

Qy        919 uHisLeuTyrProPheLeuLysGlyLeuMetGlyArgGlnAsnLeuThrProThrIleVa 939

Db       2923 CCATCTTTATCCATTCCTCAAAGGTCTTATGGGACGCCAAAACAGGACACCAACCATTGT 2982

Qy        939 lValLeuTrpSerValLeuLeuAlaSerValPheSerLeuValTrpValLysIleAsnPr 959

Db       2983 TGTCCTTTGGTCAGTGTTGTTGGCTTCTGTCTTCTCTCTTGTTGGGTTCGGATCAACCC 3042

Qy        959 cPheValAsnLysValAspAsnThrLeuValAlaGluThrCysIleSerIleAspCys 978

Db       3043 GTTTGTCAGCACCGCCGATAC ACCA TT AA AAGCT ATTTTCATTGATTGT 3100

RESULT 2
US-09-838-586-1
; Sequence 1, Application US/09838586
; Patent No. 6576818
; GENERAL INFORMATION:
;  APPLICANT: Stalker, D. et al.
;  TITLE OF INVENTION: Plant Cellulose Synthase and Promoter
;  TITLE OF INVENTION:  Sequences
;  FILE REFERENCE: 15621/02/US
;  CURRENT APPLICATION NUMBER: US/09/838,586
;  CURRENT FILING DATE:  2001-04-18
;  PRIOR APPLICATION NUMBER: 60/029,987
;  PRIOR FILING DATE: 1996-10-29
;  PRIOR APPLICATION NUMBER: 08/960,048
;  PRIOR FILING DATE: 1997-10-29
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;  LENGTH: 3328
;  TYPE: DNA
;  ORGANISM: Artificial Sequence
;  FEATURE:
;  OTHER INFORMATION: Synthetic Oligonucleotide
US-09-838-586-1

Alignment Scores:
```

Pred. No.:	0	Length:	3328
Score:	4540.00	Matches:	845
Percent Similarity:	93.27%	Conservative:	69
Best Local Similarity:	86.22%	Mismatches:	59
Query Match:	86.64%	Indels:	8
DB:	4	Gaps:	6

US-09-980-043-2 (1-978) x US-09-838-586-1 (1-3328)

```
Qy          1 MetMetGluSerGlyAlaProIleCysHisThrCysGlyGluGlnValGlyHisAspAla 20

Db        179 ATGATGGAATCTGGGGTTCCTGTTTGCCACACTTGTGGTGAACATGTTGGGTTGAATGTT 238

Qy         21 AsnGlyGluLeuPheValAlaCysHisGluCysSerTyrProMetCysLysSerCysPhe 40

Db        239 AATGGTGAACCTTTTGTGGCTTGCCATGAATGTAATTTCCCTATTTGTAAGAGTTGTTTT 298

Qy         41 GluPheGluIleAsnGluGlyArgLysValCysLeuArgCysGlySerProTyrAspGlu 60

Db        299 GAGTATGATCTTAAGGAAGGACGAAAAGCTTGCTTGCGTTGTGGTAGTCCATATGATGAA 358

Qy         61 AsnLeuLeuAspAspValGluLysLysGlySerGlyAsnGlnSerThrMetAlaSerHis 80

Db        359 AACCTGTTGGACGATGTCGAG---AAGGCCACCGGCGATCAATCGACAATGGCTGCACAT 415

Qy         81 LeuAsnAspSerGlnAspValGlyIleHisAlaArgHisIleSerSerValSerThrVal 100

Db        416 TTGAACAAGTCTCAGGATGTTGGAATTCATGCAAGACATATCAGCAGTGTGTCTACATTG 475

Qy        101 AspSerGluMetAsnAspGluTyrGlyAsnProIleTrpLysAsnArgValLysSerCys 120

Db        476 GATAGTGAAATGGCTGAAGACAATGGGAATTCGATTTGGAAGAACAGGGTGGAAAGTTGG 535

Qy        121 LysAspLysGluAsnLysLysLysLysArgSerProLysAlaGluThrGluProAlaGln 140

Db        536 AAAGAAAAGAAGAACAAGAAGAAGAAGCCTGCAACAACTAAGGTTGAAAGAGAGGCTGAA 595

Qy        141 ValProThrGluGlnGlnMetGluGluLysProSerAlaGluAlaSerGluProLeuSer 160

Db        596 ATCCCACCTGAGCAACAAATGGAAGATAAACCGGCACCGGATGCTTCCCAGCCCCTCTCG 655

Qy        161 IleValTyrProIleProArgAsnLysLeuThrProTyrArgAlaValIleIleMetArg 180

Db        656 ACTATAATTCCAATCCCGAAAAGCAGACTTGCACCATACCGAACCGTGATCATTATGCGA 715

Qy        181 LeuValIleLeuGlyLeuPhePheHisPheArgIleThrAsnProValAspSerAlaPhe 200

Db        716 TTGATCATTCTTGGTCTTTTCTTCCATTATCGAGTAACAAACCCCGTTGACAGTGCTTTT 775

Qy        201 GlyLeuTrpLeuThrSerValIleCysGluIleTrpPheAlaPheSerTrpValLeuAsp 220

Db        776 GGACTGTGGCTCACTTCAGTCATATGTGAAATCTGGTTTGCATTTTCCTGGGTGTTGGAT 835

Qy        221 GlnPheProLysTrpAsnProValAsnArgGluThrTyrIleGluArgLeuSerAlaArg 240

Db        836 CAGTTCCCTAAGTGGTATCCTGTTAACAGGGAAACATACATTGACAGACTATCTGCAAGA 895

Qy        241 TyrGluArgGluGlyGluProSerGlnLeuAlaGlyValAspPhePheValSerThrVal 260

Db        896 TATGAAAGAGAAGGTGAACCTGATGAACTTGCTGCAGTTGACTTCTTCGTGAGTACAGTG 955

Qy        261 AspProLeuLysGluProProLeuIleThrAlaAsnThrValLeuSerIleLeuAlaVal 280

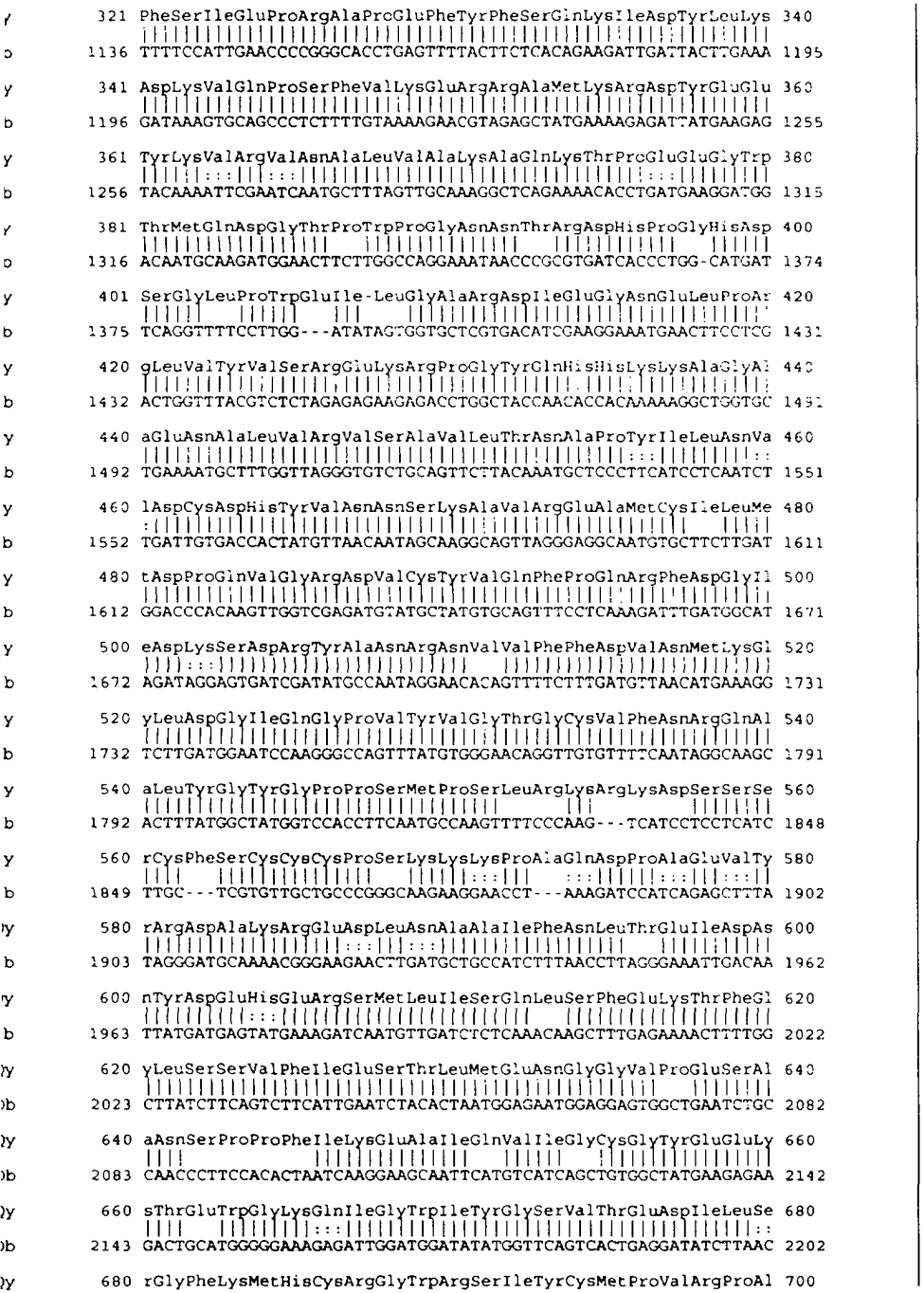
Db        956 GATCCATTGAAAGAGCCTCCATTGATTACTGCCAATACTGTGCTTTCCATCCTTGCCTTG 1015

Qy        281 AspTyrProValAspLysValSerCysTyrValSerAspAspGlyAlaAlaMetLeuSer 300

Db       1016 GACTACCCGGTGGATAAGGTCTCTTGTTATATATCTGATGATGGTGCGGCCATGCTGACA 1075

Qy        301 PheGluSerLeuValGluThrAlaGluPheAlaArgLysTrpValProPheCysLysLys 320

Db       1076 TTTGAATCTCTAGTAGAAACAGCCGACTTTGCAAGAAAGTGGGTTCCATTCTGCAAAAAA 1135
```



RESULT 3
US-09-221 013A-5
; Sequence 5, Application US/09221013A
; Patent No. 6495740
; GENERAL INFORMATION:
; APPLICANT: Arioli, Antonio
; APPLICANT: Williamson, Richard E.
; APPLICANT: Betzner, Andreas S.
; APPLICANT: Peng, Liangcai
; TITLE OF INVENTION: Manipulation of cellulose and/or Beta-1,4 glucan
; FILE REFERENCE: 96-98
; CURRENT APPLICATION NUMBER: US/09/221,013A
; CURRENT FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: PCT/AU97/00402

```
;  PRIOR FILING DATE:  1997-06-24
;  PRIOR APPLICATION NUMBER:  AU PO0699
;  PRIOR FILING DATE:  1996-06-27
;  NUMBER OF SEQ ID NOS:  37
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 5
;   LENGTH: 3603
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(3243)
US-09-221-013A-5

Alignment Scores:
Pred. No.:                0          Length:          3603
Score:                    3350.00    Matches:         648
Percent Similarity:       71.92%     Conservative:    123
Best Local Similarity:    60.45%     Mismatches:      181
Query Match:              63.93%     Indels:          120
DB:                       4          Gaps:            19

US-09-980-043-2 (1-978) x US-09-221-013A-5 (1-3603)

Qy      8 IleCysHisThrCysGlyGluGlnValGlyHisAspAlaAsnGlyGluLeuPheValAla 27
Db    112 ATATGTCAGATCTGTGGTGATGATGTTGGACTCGCTGAAACTGGAGATGTCTTTGTCGCG 171

Qy     28 CysHisGluCysSerTyrProMetCysLysSerCysPheGluPheGluIleAsnGluGly 47
Db    172 TGTAATGAATGTGCCTTCCCTGTGTGTCGGCCTTGCTATGAGTACGAGAGGAAAGATGGA 231

Qy     48 ArgLysValCysLeuArgCys ------------GlySerPro------ 57
Db    232 ACTCAGTGTTGCCCTCAATGCAAGACTAGATTCAGACGACAGGGGGAGTCCTCGTGTT 291

Qy     58 ---------TyrAspGluAsnLeuLeuAspAspValGluLysLysGlySerGlyAsnGln 74
Db    292 GAAGGAGATGAAGATGAGGATGATGTTGATGATATCGAGAATGAGTTCAATTACGCCCAG 351

Qy     75 ---SerThrMetAlaSerHisLeuAsnAspSerGlnAspValGlyIleHisAlaArgHis 93
Db    352 GGAGCTAACAAGGCGAGACACCAACGCCATGGCGAAGAGTTTTCTTCTTCCTCTAGACAT 411

Qy     94 IleSer------------------------------------SerValSerThr 99
Db    412 GAATCTCAACCAATTCCTCTTCTCACCCATGGCCATACGGTTTCTGGAGAGATTCGCACG 471

Qy    100 ValAspSerGlu-------------------------------- 103
Db    472 CCTGATACACAATCTGTGCGAACAACATCAGGTCCTTTGGGTCCTGACAGGAATGCT 531

Qy    104 MetAsnAspGluTyr------------------------------ 108
Db    532 ATTTCATCTCCATATATTGATCCACGGCAACCTGTCCCTGTAAGAATCGTGGACCCGTCA 591

Qy    109 ------------------GlyAsnProIleTrpLysAsnArgValLysSerCys 120
Db    592 AAAGACTTGAACTCTTATGGGCTTGGTAATGTTGACTGGAAAGAAAGAGTTGAAGGCTGG 651

Qy    121 LysAspLysGluAsnLys ------------------LysLysLysArgSerProLys 133
Db    652 AAGCTGAAGCAGGAGAAAAATATGTTACAGATGACTGGTAAATACCATGAAGGGAAAGGA 711

Qy    134 AlaGluThrGluProAlaGlnValProThrGluGlnGlnMetGluGluLysProSerAla 153
Db    712 GGAGAAATTGAA-----------GGGACTGGTTCCAATGGAGAAGAACTCCAAATGGCT 759

Qy    154 GluAlaSerGlu---ProLeuSerIleValTyrProIleProArgAsnLysLeuThrPro 172
Db    760 GATGATACACGTCTTCCTATGAGTCGTGTGGTGCCTATCCCATCTTCTCGCCTAACCCCT 819

Qy    173 TyrArgAlaValIleIleMetArgLeuValIleLeuGlyLeuPhePheHisPheArgIle 192
Db    820 TATCGGGTTGTGATTATTCTCCGGCTTATCATCTTGTGTTTCTTCTTGCAATATCGTACA 879

Qy    193 ThrAsnProValAspSerAlaPheGlyLeuTrpLeuThrSerValIleCysGluIleTrp 212
Db    880 ACTCACCCTGTGAAAAATGCATATCCTTTGTGGTTGACCTCGGTTATCTGTGAGATCTGG 939

Qy    213 PheAlaPheSerTrpValLeuAspGlnPheProLysTrpAsnProValAsnArgGluThr 232
Db    940 TTTGCATTTTCTTGGCTTCTTGATCAGTTTCCCAAATGGTACCCCATTAACAGGGAGACT 999

Qy    233 TyrIleGluArgLeuSerAlaArgTyrGluArgGluGlyGluProSerGlnLeuAlaGly 252
Db   1000 TATCTTGACCGTCTCGCTATAAGATATGATCGAGACGGTGAACCATCACAGCTCGTTCCT 1059

Qy    253 ValAspPhePheValSerThrValAspProLeuLysGluProProLeuIleThrAlaAsn 272
Db   1060 GTTGATGTGTTTGTTAGTACAGTGGACCCATTGAAAGAGCCTCCCCTTGTTACAGCAAAC 1119

Qy    273 ThrValLeuSerIleLeuAlaValAspTyrProValAspLysValSerCysTyrValSer 292
Db   1120 ACAGTTCTCTCGATTCTTTCTGTGGACTACCCGGTAGATAAAGTAGCCTGTTATGTTTCA 1179

Qy    293 AspAspGlyAlaAlaMetLeuSerPheGluSerLeuValGluThrAlaGluPheAlaArg 312
Db   1180 GATGATGGTTCAGCTATGCTTACCTTTGAATCCCTTTCTGAAACCGCTGAGTTTGCAAAG 1239

Qy    313 LysTrpValProPheCysLysLysPheSerIleGluProArgAlaProGluPheTyrPhe 332
Db   1240 AAATGGGTACCATTTTGCAAGAAATTCAACATTGAACCTAGGGCCCCTGAATTCTATTTT 1299

Qy    333 SerGlnLysIleAspTyrLeuLysAspLysValGlnProSerPheValLysGluArgArg 352
Db   1300 GCCCAGAAGATAGATTACTTGAAGGACAAGATCCAACCGTCTTTTGTTAAAGAGCGACGA 1359

Qy    353 AlaMetLysArgAspTyrGluGluTyrLysValArgValAsnAlaLeuValAlaLysAla 372
Db   1360 GCTATGAAGAGAGAGTATGAAGAGTTTAAAGTGAGGATAAATGCTCTTGTTGCCAAAGCA 1419

Qy    373 GlnLysThrProGluGluGlyTrpThrMetGlnAspGlyThrProTrpProGlyAsnAsn 392
Db   1420 CAGAAAATCCCTGAAGAAGGCTGGACAATGCAGGATGGTACTCCCTGGCCTGGTAACAAC 1479

Qy    393 ThrArgAspHisPro--------------------GlyHisAspSerGlyLeuProTrp 405
Db   1480 ACTAGAGATCATCCTGGAATGATACAGGTGTTCTTAGGCCATAGTGGGGGTCTG------ 1533

Qy    406 GluIleLeuGlyAlaArgAspIleGluGlyAsnGluLeuProArgLeuValTyrValSer 425
Db   1534 ------------------GATACCGATGGAAATGAGCTGCCTAGACTCATCTATGTTTCT 1575

Qy    426 ArgGluLysArgProGlyTyrGlnHisHisLysLysAlaGlyAlaGluAsnAlaLeuVal 445
Db   1576 CGTGAAAAGCGGCCTGGATTTCAACACCACAAAAAGGCTGGAGCTATGAATGCATTGATC 1635

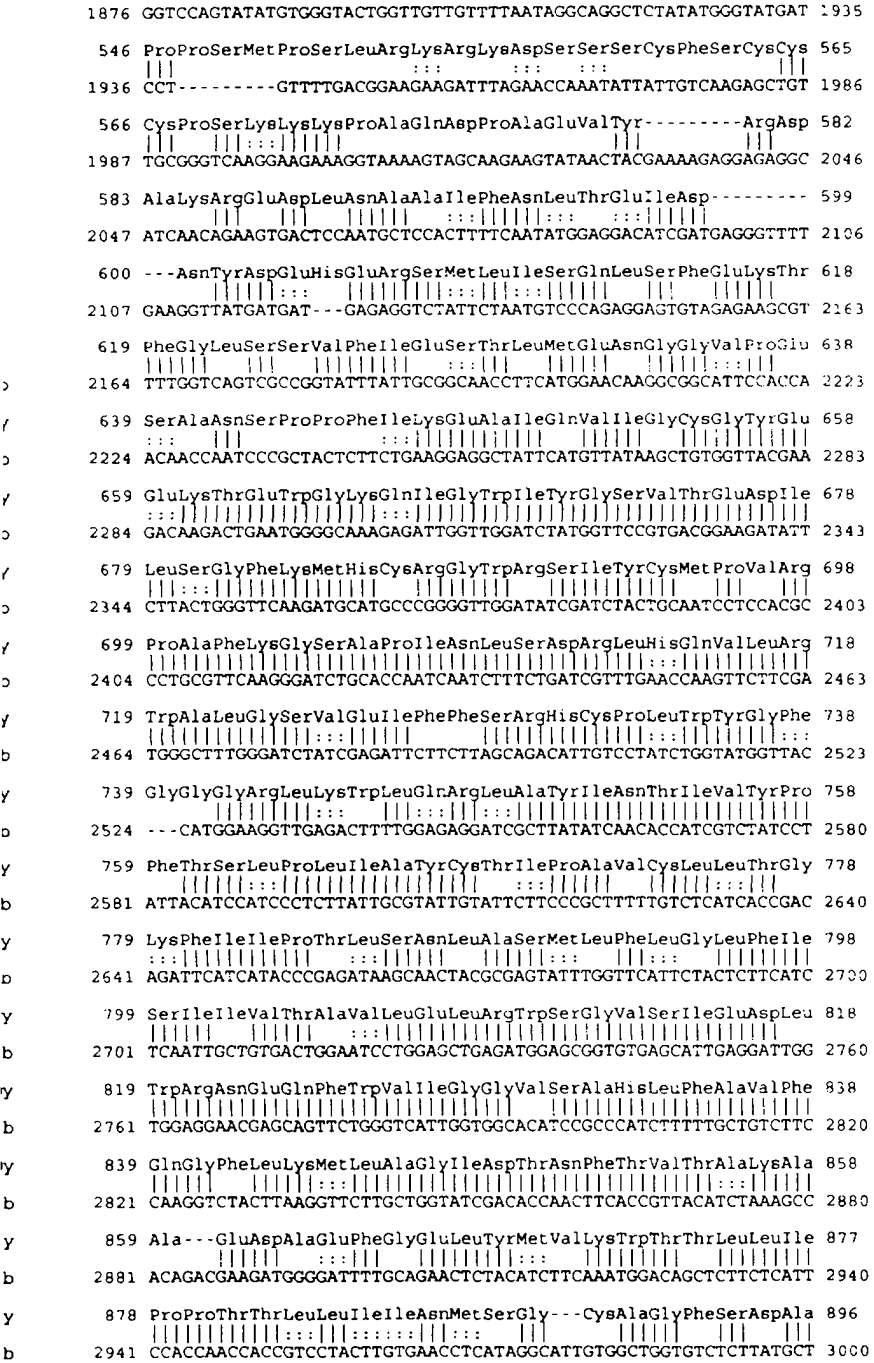
Qy    446 ArgValSerAlaValLeuThrAsnAlaProTyrIleLeuAsnValAspCysAspHisTyr 465
Db   1636 CGTGTATCTGCTGTTCTTACCAATGGAGCATATCTTTTGAACGTGGATTGTGATCATTAC 1695

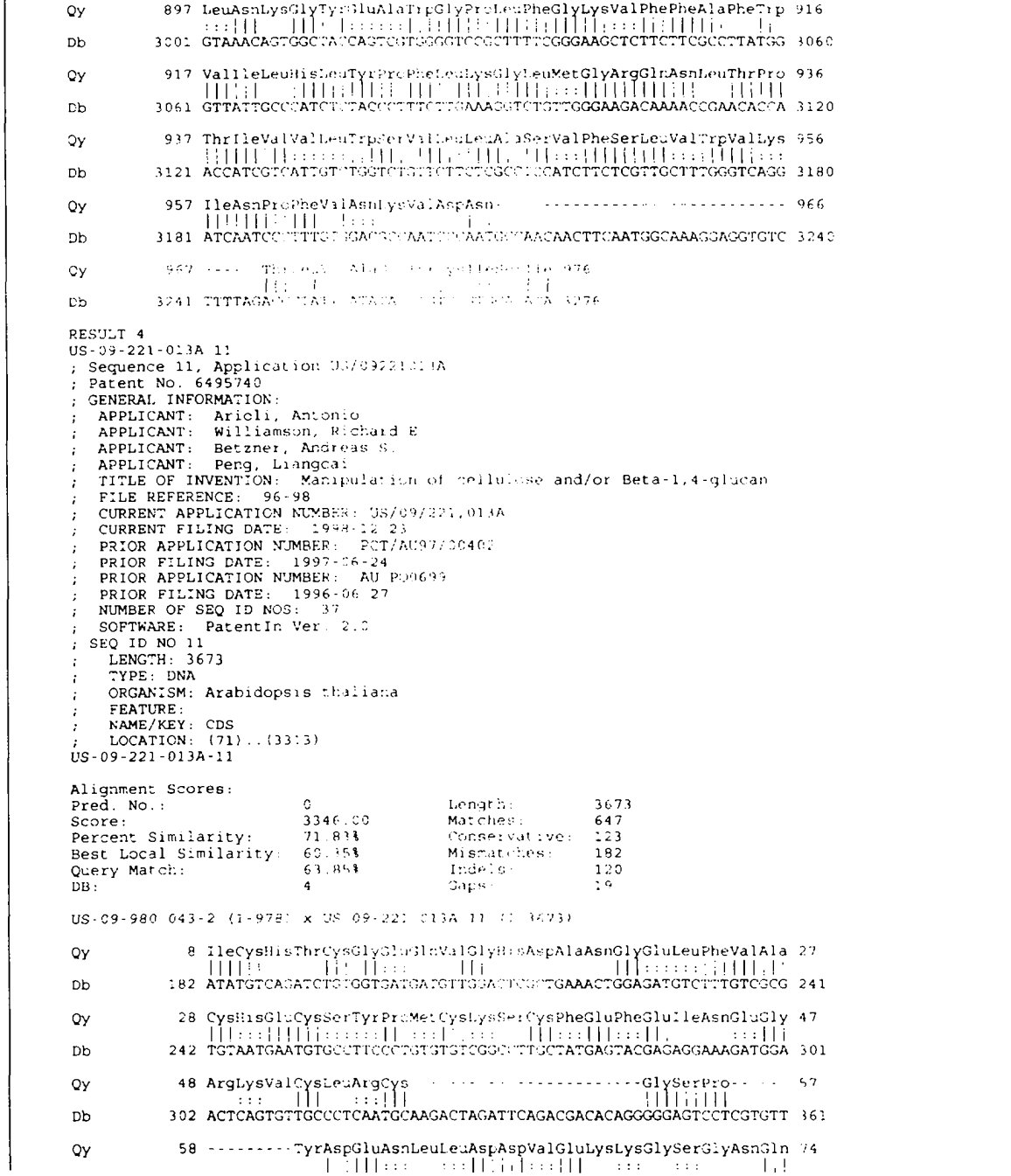
Qy    466 ValAsnAsnSerLysAlaValArgGluAlaMetCysIleLeuMetAspProGlnValGly 485
Db   1696 TTTAATAACAGTAAGGCTATTAAAGAAGCTATGTGTTTCATGATGGACCCGGCTATTGGA 1755

Qy    486 ArgAspValCysTyrValGlnPheProGlnArgPheAspGlyIleAspLysSerAspArg 505
Db   1756 AAGAAGTGCTGCTATGTCCAGTTCCCTCAACGTTTTGACGGTATTGATTTGCACGATCGA 1815

Qy    506 TyrAlaAsnArgAsnValValPhePheAspValAsnMetLysGlyLeuAspGlyIleGln 525
Db   1816 TATGCCAACAGGAATATAGTCTTTTTCGATATTAACATGAAGGGGTTGGATGGTATCCAG 1875

Qy    526 GlyProValTyrValGlyThrGlyCysValPheAsnArgGlnAlaLeuTyrGlyTyrGly 545
```





RESULT 4
US-09-221-013A 11
; Sequence 11, Application US/09/221,013A
; Patent No. 6495740
; GENERAL INFORMATION:
; APPLICANT: Aricli, Antonio
; APPLICANT: Williamson, Richard E
; APPLICANT: Betzner, Andreas S.
; APPLICANT: Peng, Liangcai
; TITLE OF INVENTION: Manipulation of cellulose and/or Beta-1,4-glucan
; FILE REFERENCE: 96-98
; CURRENT APPLICATION NUMBER: US/09/221,013A
; CURRENT FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: PCT/AU97/00407
; PRIOR FILING DATE: 1997-06-24
; PRIOR APPLICATION NUMBER: AU PO0699
; PRIOR FILING DATE: 1996-06-27
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 3673
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (71)..(3313)
US-09-221-013A-11

Alignment Scores:

Pred. No.:	0	Length:	3673
Score:	3346.00	Matches:	647
Percent Similarity:	71.83%	Conservative:	123
Best Local Similarity:	69.35%	Mismatches:	182
Query Match:	63.85%	Indels:	120
DB:	4	Gaps:	19

US-09-980-043-2 (1-978) x US-09-221-013A-11 (1-3673)



```
y        719 TrpAlaLeuGlySerValGluIlePhePheSerArgHisCysProLeuTrpTyrGlyPhe 738
             ||||||||||||||||:::|||||      |||||||||||||||:::||||||||:::
o       2534 TGGGCTTTGGGATCTATCGAGATTCTTCTTAGCAGACATTGTCCTATCTGGTATGGTTAC 2593

y        739 GlyGlyGlyArgLeuLysTrpLeuGlnArgLeuAlaTyrIleAsnThrIleValTyrPro 758
             |||||||||:::   |||:::|||:::|||||||||||||||||||||||||||||||
o       2594 ---CATGGAAGGTTGAGACTTTTGGAGAGGATCGCTTATATCAACACCATCGTCTATCCT 2650

y        759 PheThrSerLeuProLeuIleAlaTyrCysThrIleProAlaValCysLeuLeuThrGly 778
             |||||:::|||||||||||||||||||   :::||||||   ||||||:::|||
o       2651 ATTACATCCATCCCTCTTATTGCGTATTGTATTCTTCCCGCTTTTTGTCTCATCACCGAC 2710

y        779 LysPheIleIleProThrLeuSerAsnLeuAlaSerMetLeuPheLeuGlyLeuPheIle 798
             :::||||||||||||   :::||||||   ||||||:::   |||:::   |||||:|||
o       2711 AGATTCATCATACCCGAGATAAGCAACTACGCGAGTATTTGGTTCATTCTACTCTTCATC 2770

y        799 SerIleIleValThrAlaValLeuGluLeuArgTrpSerGlyValSerIleGluAspLeu 818
             ||||||   ||||||   :::||||||||||||||||||||||||||||||||||||||
o       2771 TCAATTGCTGTGACTGGAATCCTGGAGCTGAGATGGAGCGGTGTGAGCATTGAGGATTGG 2830

y        819 TrpArgAsnGluGlnPheTrpValIleGlyGlyValSerAlaHisLeuPheAlaValPhe 838
             |||||||||||||||||||||||||||||||||   |||||||||||||||||||||||
o       2831 TGGAGGAACGAGCAGTTCTGGGTCATTGGTGGCACATCCGCCCATCTTTTTGCTGTCTTC 2890

y        839 GlnGlyPheLeuLysMetLeuAlaGlyIleAspThrAsnPheThrValThrAlaLysAla 858
             ||||||   ||||||:::||||||||||||||||||||||||||||||||:::||||||
b       2891 CAAGGTCTACTTAAGGTTCTTGCTGGTATCGACACCAACTTCACCGTTACATCTAAAGCC 2950

y        859 Ala---GluAspAlaGluPheGlyGluLeuTyrMetValLysTrpThrThrLeuLeuIle 877
             |||||||   :::|||   |||||||||:::   |||||||||   |||||||||
b       2951 ACAGACGAAGATGGGGATTTTGCAGAACTCTACATCTTCAAATGGACAGCTCTTCTCATT 3010

y        878 ProProThrThrLeuLeuIleIleAsnMetSerGly---CysAlaGlyPheSerAspAla 896
             ||||||||||||:::|||:::|||:::   |||      ||||||   |||   |||
b       3011 CCACCAACCACCGTCCTACTTGTGAACCTCATAGGCATTGTGGCTGGTGTCTCTTATGCT 3070

y        897 LeuAsnLysGlyTyrGluAlaTrpGlyProLeuPheGlyLysValPhePheAlaPheTrp 916
             :::|||   ||||||:::::||||||||||||||||||||:::|||||||||   |||
b       3071 GTAAACAGTGGCTACCAGTCGTGGGGTCCGCTTTTCGGGAAGCTCTTCTTCGCCTTATGG 3130

y        917 ValIleLeuHisLeuTyrProPheLeuLysGlyLeuMetGlyArgGlnAsnLeuThrPro 936
             ||||||   ||||||||||||||||||||||||||:::||||||||||||   ||||||
b       3131 GTTATTGCCCATCTCTACCCTTTCTTGAAAGGTCTGTTGGGAAGACAAAACCGAACACCA 3190

y        937 ThrIleValValLeuTrpSerValLeuLeuAlaSerValPheSerLeuValTrpValLys 956
             |||||||||:::::|||||||||||||||||||:::||||||||:::|||||||:::
b       3191 ACCATCGTCATTGTCTGGTCTGTTCTTCTCGCCTCCATCTTCTCGTTGCTTTGGGTCAGG 3250

y        957 IleAsnProPheValAsnLysValAspAsn---------------------------- 966
             ||||||||||||||:::          |||
b       3251 ATCAATCCCTTTGTGGACGCCAATCCCAATGCCAACAACTTCAATGGCAAAGGAGGTGTC 3310

y        967 ------ThrLeuValAlaGluThrCysIleSerIle 976
             ||||||          |||:::   |||
b       3311 TTTTAGACCCTATTTATATACTTGTGTGTGCATATA 3346

ESULT 5
S-09-221-013A-9
 Sequence 9, Application US/09221013A
 Patent No. 6495740
 GENERAL INFORMATION:
  APPLICANT:  Arioli, Antonio
  APPLICANT:  Williamson, Richard E.
  APPLICANT:  Betzner, Andreas S.
  APPLICANT:  Peng, Liangcai
  TITLE OF INVENTION:  Manipulation of cellulose and/or Beta-1,4-glucan
  FILE REFERENCE:  96-98
  CURRENT APPLICATION NUMBER:  US/09/221,013A
  CURRENT FILING DATE:  1998-12-23
  PRIOR APPLICATION NUMBER:  PCT/AU97/00402
  PRIOR FILING DATE:  1997-06-24
  PRIOR APPLICATION NUMBER:  AU PO0699
  PRIOR FILING DATE:  1996-06-27

;  NUMBER OF SEQ ID NOS: 9
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 9
;   LENGTH: 3614
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (217)...(3411)
US-09-221-013A-9

Alignment Scores:
Pred. No.:                0                 Length:           3614
Score:                    332.00            Matches:          652
Percent Similarity:       71.43             Conservative:     123
Best Local Similarity:    60.43             Mismatches:       187
Query Match:              63.43             Indels:           118
DB:                       4                 Gaps:             21

US-09-980-043-2 (1-979) x US-09-221-013A-9 (217-3411)

Qy          2 MetGluSer---------GlyAsp-IleP---------------Cys 9
              ||:|||              ||   |||:::                |||
Db        217 ATGGAATCGGAAGGAGAAACCGCGGGAAAGCTGATGAAGAACATTGTTCCGCAGACTTGC 276

Qy         10 HisThrCysGlyGlnGlnValGlyHisAspAlaAsnGlyGluLeuPheValAlaCysHis 29
                |||   :::   |:|||          :::|||:::   ||||||:|||
Db        277 CAGATCTGTAGTGATAATGTTGGCAAGACTGTTGATGGAGATCGTTTTGTGGCTTGTGAT 336

Qy         30 GluCysSerTyrProMetCysLysSerCysProGluPheGluIleAsnGluGlyArgLys 49
               |||.||:::|||:::|||:::   :|:::|:::|||      :::||,   :::
Db        337 ATTTGTTTATTCCCAGTTTGTCGGCCTTCTTACGAGTATGAGAGGAAAGATGGGAATCAA 396

Qy         50 ValCysLeuArgCys-----------------------GlySerPro--------- 57
              |||   ::::|                         |||:|||||
Db        397 TCTTGTCCTCAGTGTAAAACCAGATACAACAGCCTCAAAGGTAGTCCTGCTATTCCTGGT 456

Qy         58 ---TyrAspGluAsnLeuLeuAspAspVal-------------------- 66
                 |:|||:::   :|  |||
Db        457 GATAAAGATGAGGATGGCTTAGCTGATGAAGGTACTGTTGAGTTCAACTACCCTCAGAAG 516

Qy         67 ---------------------------------------GluLysLys 69
                                                      |||:::
Db        517 GAGAAAATTTCAGATCGGATGCTTGGTTGGCATCTTACTCGTGGGAAGGGAGAGGAAATG 576

Qy         70 GlySerGlyAsnGlnSerThrMetAlaSerHis-----------LeuAsnAspSer 84
              |||                  |||:          |||
Db        577 GGGGAACCCCAGTATGATAAAGAGGTCTCTCACAATCATCTTCCTCGTCTCACGAGCAGA 636

Qy         85 GlnAspVal-------GlyIleHisAlaArgHisIleSerSerValSerThr 99
              ||||||                  :::|||:    ||||:|:
Db        637 CAAGATACTTCAGGAGAGTTTTCTGCTGCCTCACCTGAACGCCTCTCTGTATCTTCTACT 696

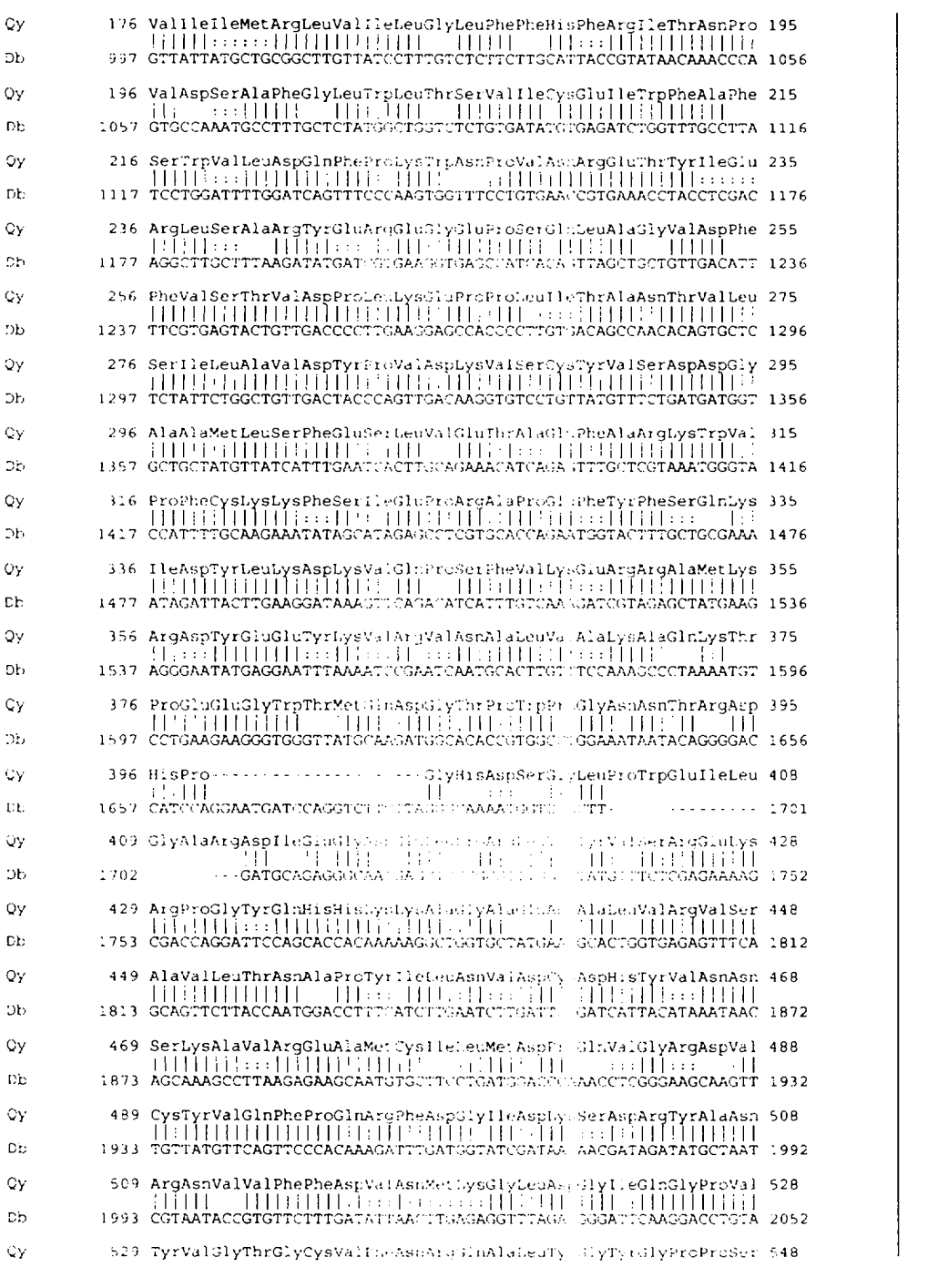
Qy        100 Val-------------AspSerGluMetAsnAspGlu---------- 107
              :::                 || ::::::|||
Db        697 ATCGCTGGTGGAAATGCTTTCCCTATTCATCAGATGTCAATCAATCACCAAATAGAAGG 756

Qy        108 -------------TyrGlyAsnProIleTrpLysAsnArgValLysSerCysLys 121
                          ||:||      |:|||   |||||:          :|.
Db        757 ATTGTGGATCCTGTTGAATCGGGAATGTAGCTTGGAAGGAGAGAGTTGATGGCTGGAAA 816

Qy        122 AspLysGluAsnLysLysLys---LysArgSerProLysAlaGluThrGluProAlaGln 140
              |||:::   |||          |||   :::|||   :::|||
Db        817 ATGAAGCAAGAGAAGAATACTGGTCCTGTCAGCACGCAGGCTGCTTCTGAAAGAGGTGGA 876

Qy        141 ValProThrGluGlnGlnMet-------GluGluLysProSerAlaGluAla 155
              |||      :::                |||          :::   ||||:
Db        877 GTAGATATTGATGCCAGCACAGATATCCTAGCAGATGAGGCTCTGCTGAATGACGAAGCG 936

Qy        156 SerGluProLeuSerIleValTyrProIleProArgAsnLysLeuThrProTyrArgAla 175
              :::   ||||||           ||||||   :::::::::   |||||||||
Db        937 AGGCAGCTTCTGTCAAGGAAAGTTTCAATTCCTTCATCACGGATCAATCCTTACAGAATG 996
```


Qy 176 ValIleIleMetArgLeuValIleLeuGlyLeuPhePheHisPheArgIleThrAsnPro 195
Db 997 GTTATTATGCTGCGGCTTGTTATCCTTTGTCTCTTCTTGCATTACCGTATAACAAACCCA 1056
Qy 196 ValAspSerAlaPheGlyLeuTrpLeuThrSerValIleCysGluIleTrpPheAlaPhe 215
Db 1057 GTGCCAAATGCCTTTGCTCTATGGCTGGTCTCTGTGATATGTGAGATCTGGTTTGCCTTA 1116
Qy 216 SerTrpValLeuAspGlnPheProLysTrpAsnProValAsnArgGluThrTyrIleGlu 235
Db 1117 TCCTGGATTTTGGATCAGTTTCCCAAGTGGTTTCCTGTGAACCGTGAAACCTACCTCGAC 1176
Qy 236 ArgLeuSerAlaArgTyrGluArgGluGlyGluProSerGlnLeuAlaGlyValAspPhe 255
Db 1177 AGGCTTGCTTTAAGATATGATCTTGAAGGTGAGCCATCACATTAGCTGCTGTTGACATT 1236
Qy 256 PheValSerThrValAspProLeuLysGluProProLeuIleThrAlaAsnThrValLeu 275
Db 1237 TTCGTGAGTACTGTTGACCCCTTGAAGGAGCCACCCCTTGTGACAGCCAACACAGTGCTC 1296
Qy 276 SerIleLeuAlaValAspTyrProValAspLysValSerCysTyrValSerAspAspGly 295
Db 1297 TCTATTCTGGCTGTTGACTACCCAGTTGACAAGGTGTCCTGTTATGTTTCTGATGATGGT 1356
Qy 296 AlaAlaMetLeuSerPheGluSerLeuValGluThrAlaGluPheAlaArgLysTrpVal 315
Db 1357 GCTGCTATGTTATCATTTGAATCACTTGCAGAAACATCAGAATTTGCTCGTAAATGGGTA 1416
Qy 316 ProPheCysLysLysPheSerIleGluProArgAlaProGluPheTyrPheSerGlnLys 335
Db 1417 CCATTTTGCAAGAAATATAGCATAGAGCCTCGTGCACCAGAATGGTACTTTGCTGCGAAA 1476
Qy 336 IleAspTyrLeuLysAspLysValGluProSerPheValLysGluArgArgAlaMetLys 355
Db 1477 ATAGATTACTTGAAGGATAAAGTTGAGACATCATTTGTCAAAGATCGTAGAGCTATGAAG 1536
Qy 356 ArgAspTyrGluGluTyrLysValArgValAsnAlaLeuValAlaLysAlaGlnLysThr 375
Db 1537 AGGGAATATGAGGAATTTAAAATTCGAATCAATGCACTTGTTTCCAAAGCCCTAAAATGT 1596
Qy 376 ProGluGluGlyTrpThrMetGlnAspGlyThrProTrpProGlyAsnAsnThrArgAsp 395
Db 1597 CCTGAAGAAGGGTGGGTTATGCAAGATGGCACACCGTGGCCTGGAAATAATACAGGGGAC 1656
Qy 396 HisPro---------- ---GlyHisAspSerGlyLeuProTrpGluIleLeu 408
Db 1657 CATCCAGGAATGATCCAGGTCT TAAAATGGTC TT- --------- 1701
Qy 409 GlyAlaArgAspIleGlnGly ArgValSerArgGluLys 428
Db 1702 ---GATGCAGAGGCAA GCATGTTCTCGAGAAAAG 1752
Qy 429 ArgProGlyTyrGlnHisHisLysLysAlaGlyAlaIleAlaLeuValArgValSer 448
Db 1753 CGACCAGGATTCCAGCACCACAAAAAGGCTGGTGCTATGAAGCACTGGTGAGAGTTTCA 1812
Qy 449 AlaValLeuThrAsnAlaProTyrIleLeuAsnValAspCysAspHisTyrValAsnAsn 468
Db 1813 GCAGTTCTTACCAATGGACCTTTTATCTTGAATCTTGATTTGATCATTACATAAATAAC 1872
Qy 469 SerLysAlaValArgGluAlaMetCysIleLeuMetAspProGlnValGlyArgAspVal 488
Db 1873 AGCAAAGCCTTAAGAGAAGCAATGTGCTTCCTGATGGAGCCAACCTCGGGAAGCAAGTT 1932
Qy 489 CysTyrValGlnPheProGlnArgPheAspGlyIleAspLysSerAspArgTyrAlaAsn 508
Db 1933 TGTTATGTTCAGTTCCCACAAAGATTTGATGGTATCGATAAAACGATAGATATGCTAAT 1992
Qy 509 ArgAsnValValPhePheAspValAsnMetLysGlyLeuAspGlyIleGlnGlyProVal 528
Db 1993 CGTAATACCGTGTTCTTTGATATTAACTTGAGAGGTTTAGAGGGATTCAAGGACCTGTA 2052
Qy 529 TyrValGlyThrGlyCysValPheAsnArgGlnAlaLeuTyrGlyTyrGlyProProSer 548
Db 2053 TATGTCGGAACTGGATGTGTTTTCAACAGAACAGCATTATACGGTTATGAACCTCCAATA 2112
Qy 549 MetProSerLeuArgLysArgLysAspSerSerSerCysPheSerCysCysCysProSer 568
Db 2113 AAAGTA---------AAACACAAGAAGCCAAGTCTTTTATCTAAGCTCTGTGGTGGATCA 2163
Qy 569 LysLysLys------ProAlaGlnAspProAlaGluValTyrArgAspAlaLysArgGlu 586
Db 2164 AGAAAGAAGAATTCCAAAGCTAAGAAAGAGTCGGACAAAAAGAAATCAGGCAGGCATACT 2223
Qy 587 AspLeuAsnAlaAlaIlePheAsnLeuThrGluIleAsp----------Asn 600
Db 2224 GACTCAACTGTTCCTGTATTCAACCTCGATGACATAGAAGAGGGAGTTGAAGGTGCTGGT 2283
Qy 601 TyrAspGluHisGluArgSerMetLeuIleSerGlnLeuSerPheGluLysThrPheGly 620
Db 2284 TTTGATGAT---GAAAAGGCGCTCTTAATGTCGCAAATGAGCCTGGAGAAGCGATTTGGA 2340
Qy 621 LeuSerSerValPheIleGluSerThrLeuMetGluAsnGlyGlyValProGluSerAla 640
Db 2341 CAGTCTGCTGTTTTTGTTGCTTCTACCCTAATGGAAAATGGTGGTGTTCCTCCTTCAGCA 2400
Qy 641 AsnSerProProPheIleLysGluAlaIleGlnValIleGlyCysGlyTyrGluGluLys 660
Db 2401 ACTCCAGAAAACTTTCTCAAAGAGGCTATCCATGTCATTAGTTGTGGTTATGAGGATAAG 2460
Qy 661 ThrGluTrpGlyLysGlnIleGlyTrpIleTyrGlySerValThrGluAspIleLeuSer 680
Db 2461 TCAGATTGGGGAATGGAGATTGGATGGATCTATGGTTCTGTGACAGAAGATATTCTGACT 2520
Qy 681 GlyPheLysMetHisCysArgGlyTrpArgSerIleTyrCysMetProValArgProAla 700
Db 2521 GGGTTCAAAATGCATGCCCGTGGATGGCGATCCATTTACTGCATGCCTAAGCTTCCAGCT 2580
Qy 701 PheLysGlySerAlaProIleAsnLeuSerAspArgLeuHisGlnValLeuArgTrpAla 720
Db 2581 TTCAAGGGTTCTGCTCCTATCAATCTTTCAGATCGTCTGAACCAAGTGCTGAGGTGGGCT 2640
Qy 721 LeuGlySerValGluIlePhePheSerArgHisCysProLeuTrpTyrGlyPheGlyGly 740
Db 2641 TTAGGTTCAGTTGAGATTCTCTTCAGTCGGCATTGTCCTATATGGTATGGTTAC---AAT 2697
Qy 741 GlyArgLeuLysTrpLeuGlnArgLeuAlaTyrIleAsnThrIleValTyrProPheThr 760
Db 2698 GGGAGGCTAAAATTTCTTGAGAGGTTTGCGTATGTGAACACCACCATCTACCCTATCACC 2757
Qy 761 SerLeuProLeuIleAlaTyrCysThrIleProAlaValCysLeuLeuThrGlyLysPhe 780
Db 2758 TCCATTCCTCTTCTCATGTATTGTACATTGCTAGCCGTTTGTCTCTTCACCAACCAGTTT 2817
Qy 781 IleIleProThrLeuSerAsnLeuAlaSerMetLeuPheLeuGlyLeuPheIleSerIle 800
Db 2818 ATTATTCCTCAGATTAGTAACATTGCAAGTATATGGTTTCTGTCTCTCTTTCTCTCCATT 2877
Qy 801 IleValThrAlaValLeuGluLeuArgTrpSerGlyValSerIleGluAspLeuTrpArg 820
Db 2878 TTCGCCACGGGTATACTAGAAATGAGGTGGAGTGGCGTAGGCATAGACGAATGGTGGAGA 2937
Qy 821 AsnGluGlnPheTrpValIleGlyGlyValSerAlaHisLeuPheAlaValPheGlnGly 840
Db 2938 AACGAGCAGTTTTGGGTCATTGGTGGAGTATCCGCTCATTTATTCGCTGTGTTTCAAGGT 2997
Qy 841 PheLeuLysMetLeuAlaGlyIleAspThrAsnPheThrValThrAlaLysAlaAla--- 859
Db 2998 ATCCTCAAAGTCCTTGCCGGTATTGACACAAACTTCACAGTTACCTCAAAAGCTTCAGAT 3057
Qy 860 GluAspAlaGluPheGlyGluLeuTyrMetValLysTrpThrThrLeuLeuIleProPro 879
Db 3058 GAAGACGGAGACTTTGCTGAGCTCTACTTGTTCAAATGGACAACACTTCTGATTCCGCCA 3117
Qy 880 ThrThrLeuLeuIleIleAsnMetSerGly---CysAlaGlyPheSerAspAlaLeuAsn 898

```
)      3118 ACGACGCTGCTCATTGTAAACTTAGTGGGAGTTGTTGCAGGAGTCTCTTATGCTATCAAC 3177

/       899 LysGlyTyrGluAlaTrpGlyProLeuPheGlyLysValPhePheAlaPheTrpValIle 918
)      3178 AGTGGATACCAATCATGGGGACCACTCTTTGGTAAGTTGTTCTTTGCCTTCTGGGTGATT 3237

/       919 LeuHisLeuTyrProPheLeuLysGlyLeuMetGlyArgGlnAsnLeuThrProThrIle 938
)      3238 GTTCACTTGTACCCTTTCCTCAAGGGTTTGATGGGTCGACAGAACCGGACTCCTACCATT 3297

/       939 ValValLeuTrpSerValLeuLeuAlaSerValPheSerLeuValTrpValLysIleAsn 958
)      3298 GTTGTGGTCTGGTCTGTTCTCTTGGCTTCTATCTTCTCGTTGTTGTGGGTTAGGATTGAT 3357

y       959 ProPheValAsnLysValAspAsnThrLeuValAlaGluThrCysIleSerIleAspCys 978
b      3358 CCCTTCACTAGCCGAGTCACTGGCCCGGACATTCTGGAA -----TGTGGAATCAACTGT 3411
```

ESULT 6
S-09-221-013A-7
 Sequence 7, Application US/09221013A
 Patent No. 6495740
 GENERAL INFORMATION:
  APPLICANT: Arioli, Antonio
  APPLICANT: Williamson, Richard E.
  APPLICANT: Betzner, Andreas S.
  APPLICANT: Peng, Liangcai
  TITLE OF INVENTION: *Manipulation of cellulose and/or Beta 1,4-glucan*
  FILE REFERENCE: 96-98
  CURRENT APPLICATION NUMBER: US/09/221,013A
  CURRENT FILING DATE: 1998-12-23
  PRIOR APPLICATION NUMBER: PCT/AU97/00402
  PRIOR FILING DATE: 1997-06-24
  PRIOR APPLICATION NUMBER: AU PO0699
  PRIOR FILING DATE: 1996-06-27
  NUMBER OF SEQ ID NOS: 37
  SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO 7
   LENGTH: 3828
   TYPE: DNA
   ORGANISM: Arabidopsis thaliana
   FEATURE:
   NAME/KEY: CDS
   LOCATION: (239)..(3490)
S-09-221-013A-7

ılignment Scores:			
red. No.:	0	Length:	3828
core:	3242.00	Matches:	621
ercent Similarity:	72.85%	Conservative:	157
est Local Similarity:	58.15%	Mismatches:	186
uery Match:	61.87%	Indels:	108
B:	4	Gaps:	18

```
JS-09-980-043-2 (1-978) x US-09-221-013A-7 (1-3828)

)y        1 MetMetGluSerGlyAlaProIleCysHisThrCysGlyGluGlnValGlyHisAspAla 20
)b      329 GTACAAGAACTGAGTGGGCAAACATGTCAAATCTGTGGAGATGAAATCGAATTAACGGTT 388

Qy       21 AsnGlyGluLeuPheValAlaCysHisGluCysSerTyrProMetCysLysSerCysPhe 40
)b      389 AGCAGTGAGCTCTTTGTTGCTTGCAACGAATGCGCATTCCCGGTTTGTAGACCATGCTAT 448

Qy       41 GluPheGluIleAsnGluGlyArgLysValCysLeuArgCys------------------ 54
Db      449 GAGTATGAACGTAGAGAAGGAAATCAAGCTTGTCCTCAGTGCAAAACTCGATACAAAAGG 508

Qy       55 ------GlySerPro------------------TyrAspGluAsnLeuLeuAspAspVal 66
Db      509 ATTAAAGGTAGTCCACGGGTTGATGGAGATGATGAAGAAGAAGACATTGATGATCTT 568

Qy       67 GluLysLys---------------------GlySerGlyAsnGlnSerThrMetAlaSer 79

Db      569 GAGTATGATTTTGATATGGGATGGACCCTGAACATGCCGCTGAAGCCGCACTCTCTTCA 628

Qy       80 HisLeuAsn---------------AspSer------------------------------ 84
Db      629 CGCCTTAACACCGGTTGTGGTGGATGGATTCAGCTCCACCTGGCTCTCAGATTCCTCTT 688

Qy       85 ---------------GlnAspValGlyIleHisAla---ArgHis------IleSerSer 96
Db      689 TTGACTTATTGTGATGAAGATGCTGATATGTATTCTGATCGTCATGCTCTTATCGTGCCT 748

Qy       97 ValSerThr ----                ---- ValAspSerGluMetAsn 105
Db      749 CCTTCAACGGGATAAGGAATGGCGTCTATCCTGCACCGTTTACAGATTCTTCTGCACCT 808

Qy      106 AspGlu --                        -------------TyrGlyAsnPro 111
Db      809 CCACAGGAGATAAAAAAGTTCAAAGAAAATGCGGAATATGGTTATGGAAGTGTT 868

Qy      112 IleTrpLysAsnArgGlyValGlyGluLysGluAsnLysLysLysLysArgSe 131
Db      869 GCTTGGAAGAGAGGAAAGAAGATTTCAAGAAGACAAGGCGAAAAGCTTCAAGTCATT 928

Qy      131 rProLysAlaGluThrGluProAlaGlnValProThrGluGlnGlnMetGluGluLysPr 151
Db      929 AAGCATGAAGGAGGAAACAATGGTCGAGGTTCAAT-GATGACGACGAACTAGATGATCC 987

Qy      151 oSerAla ---    --- GluAlaSerGluProLeuSerIleValTyrProIleProAr 167
Db      988 TGACATGCCTATGATGGATGAAGGAAGACAACCTCTCTCAAGAAAGCTACCTATTCGTTC 1047

Qy      167 gAsnLysLeuThrProTyrArgAlaValIleIleMetArgLeuValIleLeuGlyLeuPh 187
Db     1048 AAGCAGAATAAATCCTTACAGGATGTTAATTCTGTGTCGCCTCGCGATTCTTGGTCTTTT 1107

Qy      187 ePheHisPheArgIleThrAsnProValAspSerAlaPheGlyLeuTrpLeuThrSerVa 207
Db     1108 CTTTCATTATAGAATTCTCCATCCAGTCAATGATGCATATGGATTATGGTTAACGTCAGT 1167

Qy      207 lIleCysGluIleTrpPheAlaPheSerTrpValLeuAspGlnPheProLysTrpAsnPr 227
Db     1168 TATATGCGAAATATTGTTTGCAGTGTCTTGGATTCTTGATCAATTCCCCAAATGGTATCC 1227

Qy      227 oValAsnArgGluThrTyrIleGluArgLeuSerAlaArgTyrGluArgGluGlyGluPr 247
Db     1228 TATAGAACGTGAAACATACCTCGATAGACTCTCTCTCAGGTACGAGAAGGAAGGAAAACC 1287

Qy      247 oSerGlnLeuAlaGlyValAspPhePheValSerThrValAspProLeuLysGluProPr 267
Db     1288 GTCAGGATTAGCACCTGTTGATGTTTTTGTTAGTACAGTGGATCCGTTGAAAGAGCCCCC 1347

Qy      267 oLeuIleThrAlaAsnThrValLeuSerIleLeuAlaValAspTyrProValAspLysVa 287
Db     1348 CTTGATTACAGCAAACACAGTTCTTTCCATTCTAGCAGTTGATTATCCTGTGGATAAGGT 1407

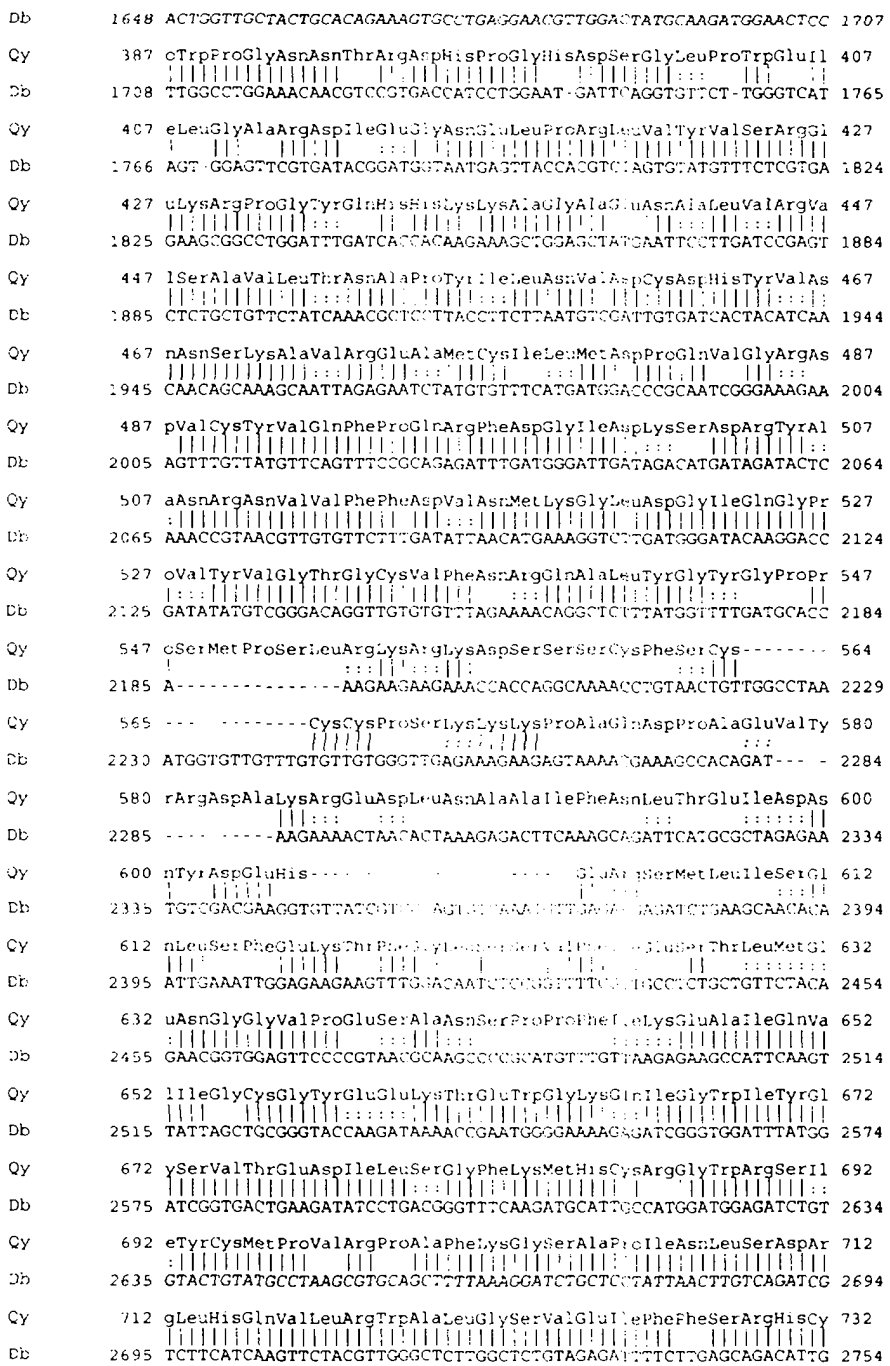
Qy      287 lSerCysTyrValSerAspAspGlyAlaAlaMetLeuSerPheGluSerLeuValGluTh 307
Db     1408 TGCGTGTTATGTATCAAACAATGGTGCAGCTATGCTTACATTTGAAGCTCTCTCTGATAC 1467

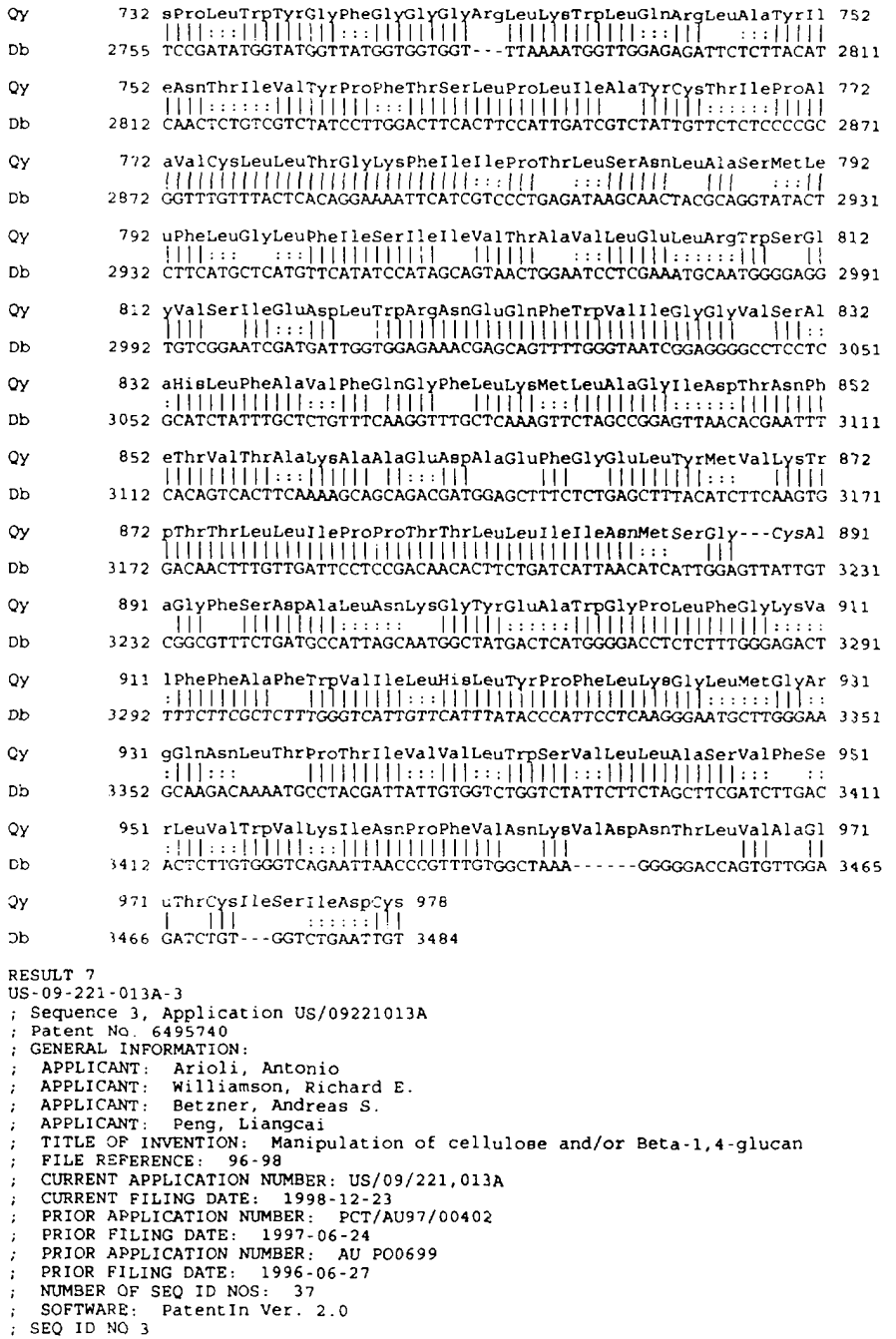
Qy      307 rAlaGluPheAlaArgLysTrpValProPheCysLysLysPheSerIleGluProArgAl 327
Db     1468 AGCTGATTTTGCTAGAAAATGGGTTCCTTTTTGTAAGAAGTTTAATATCGAGCCACGAGC 1527

Qy      327 aProGluPheTyrPheSerGlnLysIleAspTyrLeuLysAspLysValGlnProSerPh 347
Db     1528 TCCTGAGTGGTATTTTTCTCAGAAGATGGATTACCTGAAGAACAAAGTTCATCCTGCTTT 1587

Qy      347 eValLysGluArgArgAlaMetLysArgAspTyrGluGluTyrLysValArgValAsnAl 367
Db     1588 TGTCAGGGAACGTCGTGCTATGAAGAGAGATTATGAAGAGTTTAAAGTGAAGATAAATGC 1647

Qy      367 aLeuValAlaLysAlaGlnLysThrProGluGluGlyTrpThrMetGlnAspGlyThrPr 387
```





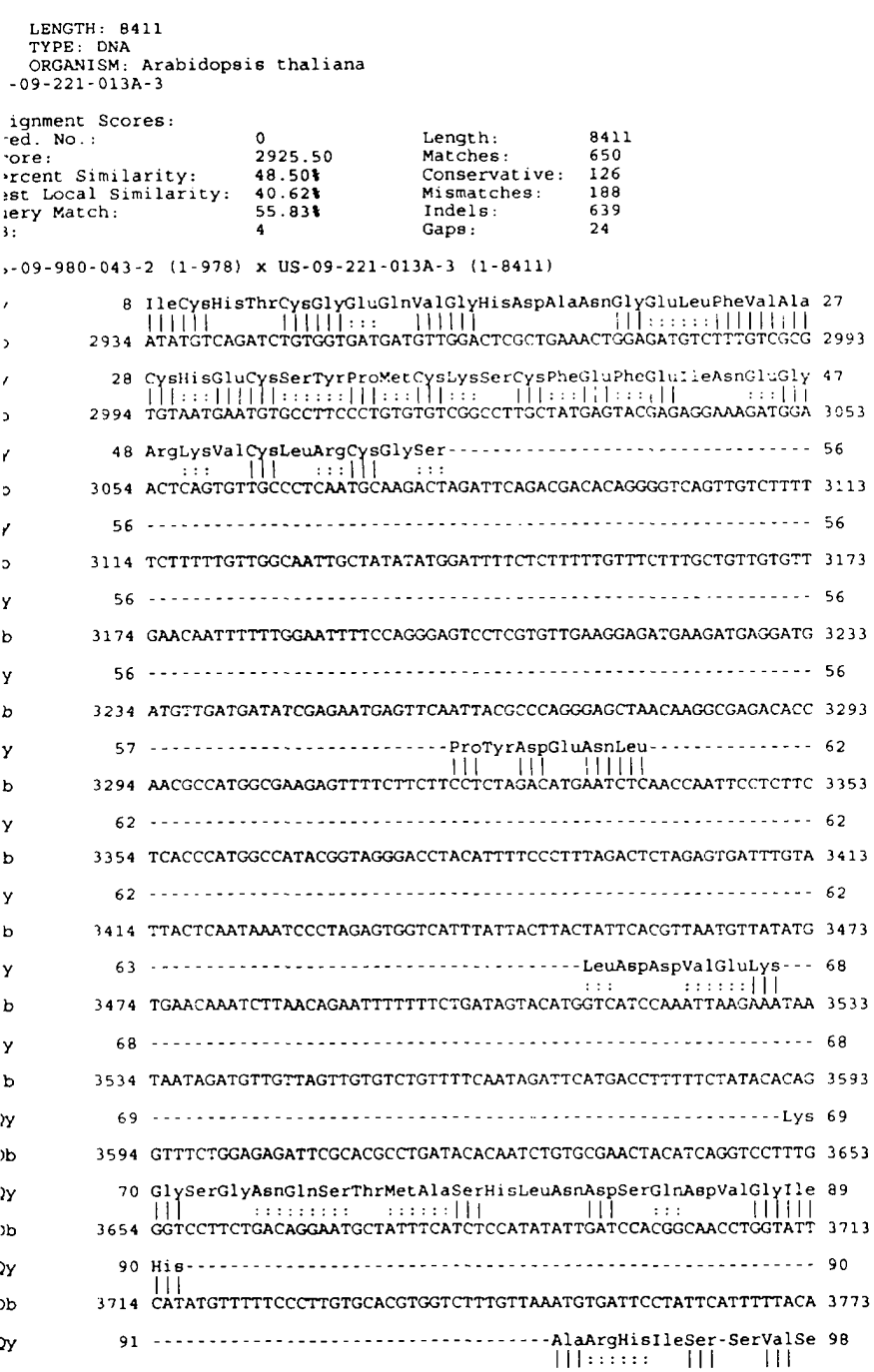
RESULT 7
US-09-221-013A-3
; Sequence 3, Application US/09221013A
; Patent No. 6495740
; GENERAL INFORMATION:
; APPLICANT: Arioli, Antonio
; APPLICANT: Williamson, Richard E.
; APPLICANT: Betzner, Andreas S.
; APPLICANT: Peng, Liangcai
; TITLE OF INVENTION: Manipulation of cellulose and/or Beta-1,4-glucan
; FILE REFERENCE: 96-98
; CURRENT APPLICATION NUMBER: US/09/221,013A
; CURRENT FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: PCT/AU97/00402
; PRIOR FILING DATE: 1997-06-24
; PRIOR APPLICATION NUMBER: AU PO0699
; PRIOR FILING DATE: 1996-06-27
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3

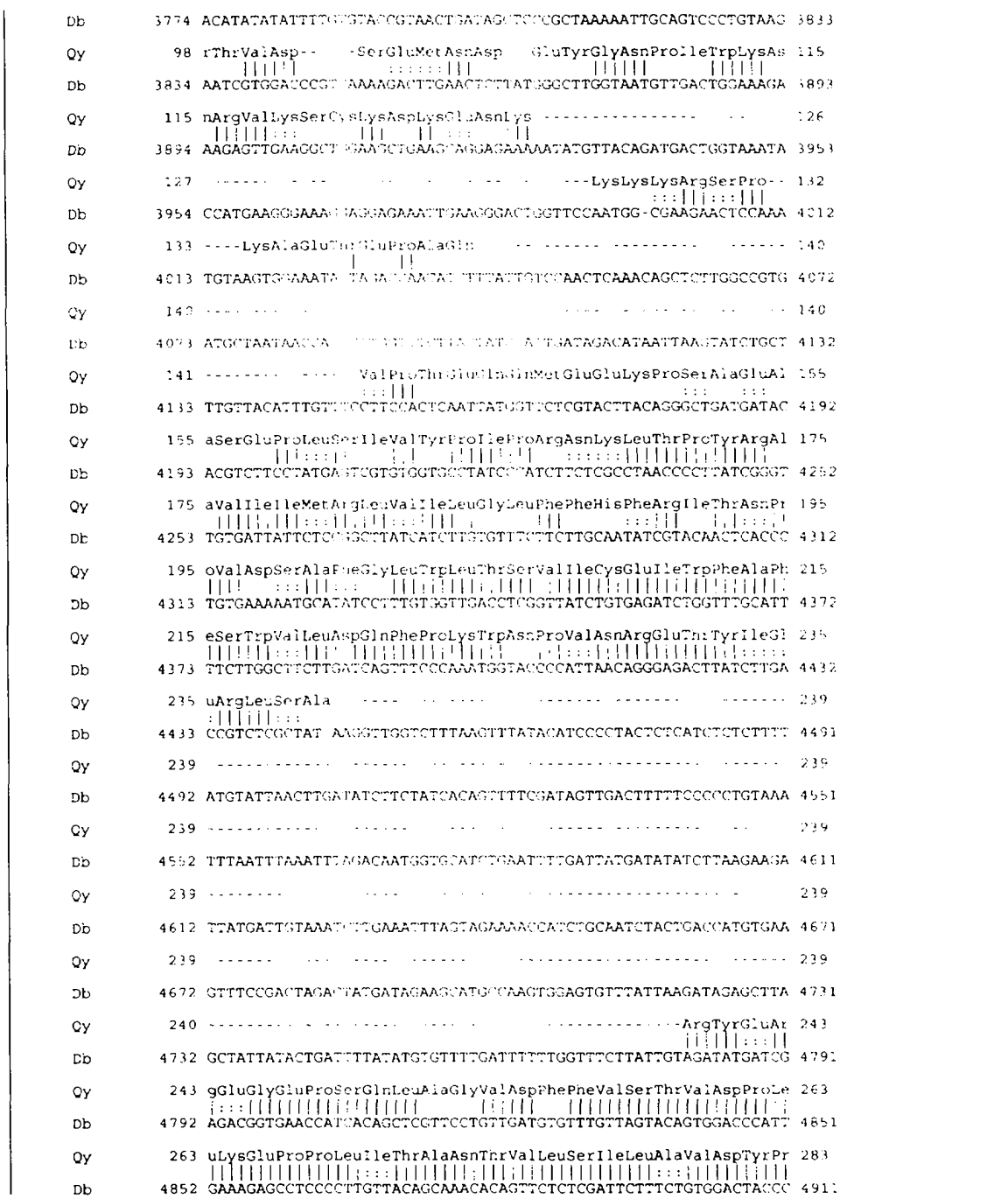
LENGTH: 8411
TYPE: DNA
ORGANISM: Arabidopsis thaliana
-09-221-013A-3

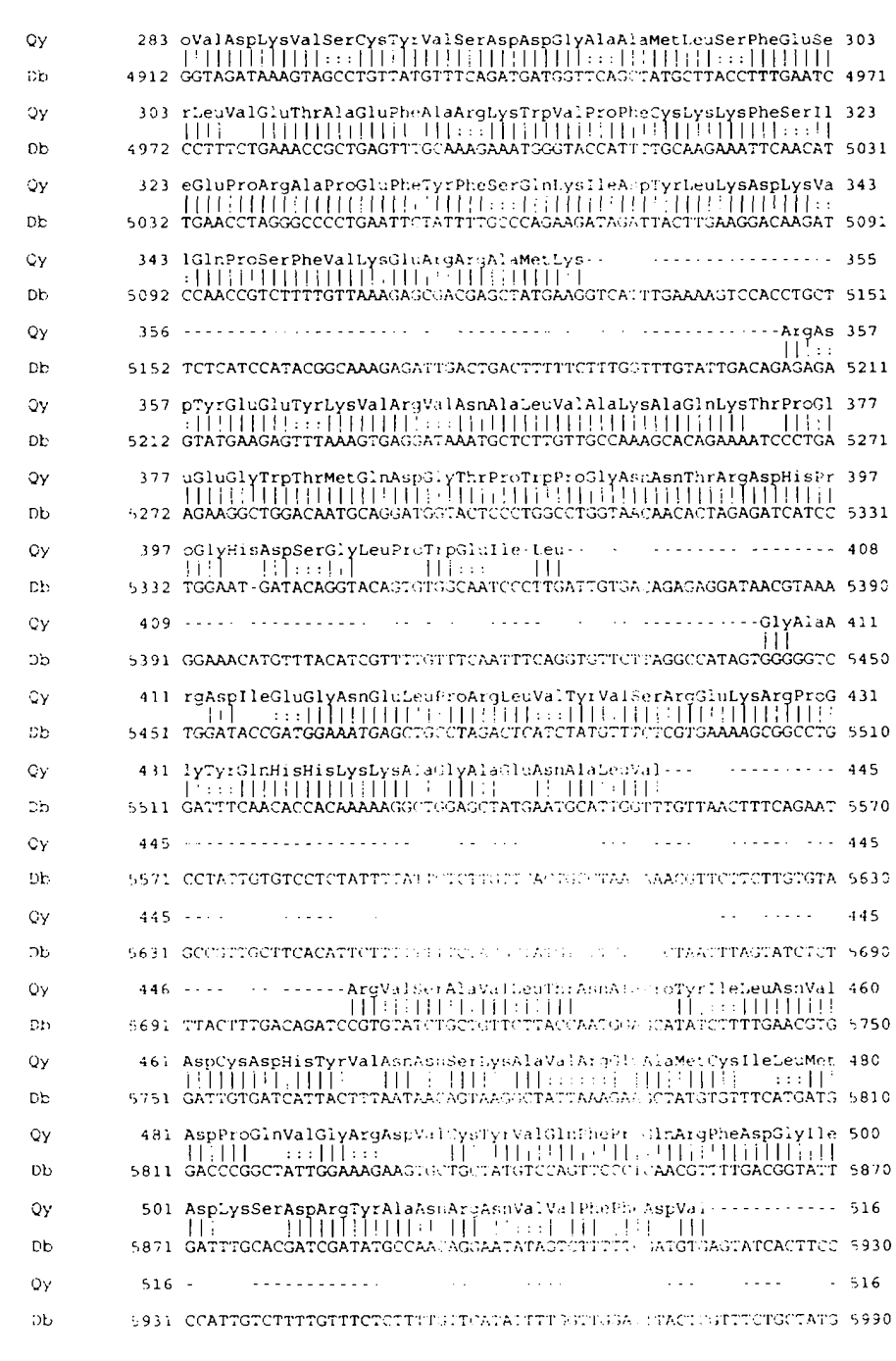
ignment Scores:

ed. No.:	0	Length:	8411
core:	2925.50	Matches:	650
ercent Similarity:	48.50%	Conservative:	126
st Local Similarity:	40.62%	Mismatches:	188
ery Match:	55.83%	Indels:	639
:	4	Gaps:	24

-09-980-043-2 (1-978) x US-09-221-013A-3 (1-8411)









```
                                                  :::||||||   ||||||:::   |||:
b       7056 AATGCATTTTTTTTGTCATCATCATTGTTGCAGATAAGCAACTACGCGAGTATTTGGTTCA 7115

y        794 euGlyLeuPheIleSerIleIleValThrAlaValLeuGluLeuArgTrpSerGlyValS 814
             ::  ||||||||||||||||   ||||||  :::||||||||||||||||||||||||||| 
b       7116 TTCTACTCTTCATCTCAATTGCTGTGACTGGAATCCTGGAGCTGAGATGGAGCGGTGTGA 7175

y        814 erIleGluAspLeuTrpArgAsnGluGlnPheTrpValIleGlyGlyValSerAlaHisL 834
             ||||||||||   ||||||||||||||||||||||||||||||||   |||||||||| 
b       7176 GCATTGAGGATTGGTGGAGGAACGAGCAGTTCTGGGTCATTGGTGGCACATCCGCCCATC 7235

y        834 euPheAlaValPheGlnGlyPheLeuLysMetLeuAlaGlyIleAspThrAsnPheThrV 854
             ||||||||||||||||||||   ||||||:::|||||||||||||||||||||||||||| 
b       7236 TTTTTGCTGTCTTCCAAGGTCTACTTAAGGTTCTTGCTGGTATCGACACCAACTTCACCG 7295

y        854 alThrAlaLysAlaAla---GluAspAlaGluPheGlyGluLeuTyrMetValLysTrpT 873
             |||||:::||||||      ||||||  :::|||  |||||||||:::  |||||||| 
b       7296 TTACATCTAAAGCCACAGACGAAGATGGGGATTTTGCAGAACTCTACATCTTCAAATGGA 7355

y        873 hrThrLeuLeuIleProProThrThrLeuLeuIleIleAsnMetSerGly---CysAlaG 892
             ||  ||||||||||||||||||||:::|||:::::::||:::  |||     |||:  
b       7356 CAGCTCTTCTCATTCCACCAACCACCGTCCTACTTGTGAACCTCATAGGCATTGTGGCTG 7415

y        892 lyPheSerAspAlaLeuAsnLysGlyTyrGluAlaTrpGlyProLeuPheGlyLysValP 912
             ||  |||  |||:::|||  ||||||:::::||||||||||||||||||||:::| 
b       7416 GTGTCTCTTATGCTGTAAACAGTGGCTACCAGTCGTGGGGTCCGCTTTTCGGGAAGCTCT 7475

y        912 hePheAlaPheTrpValIleLeuHisLeuTyrProPheLeuLysGlyLeuMetGlyArgG 932
             |||||||  |||||||||  ||||||||||||||||||||||||:::|||||||| 
b       7476 TCTTCGCCTTATGGGTTATTGCCCATCTCTACCCTTTCTTGAAAGGTCTGTTGGGAAGAC 7535

y        932 lnAsnLeuThrProThrIleValValLeuTrpSerValLeuLeuAlaSerValPheSerL 952
             |||||  ||||||||||||||||:::::|||||||||||||||||||||:::|||||| 
b       7536 AAAACCGAACACCAACCATCGTCATTGTCTGGTCTGTTCTTCTCGCCTCCATCTTCTCGT 7595

y        952 euValTrpValLysIleAsnProPheValAsnLysValAspAsn---------------- 966
             ||:::||||||:::|||||||||||||||:::          |||
b       7596 TGCTTTGGGTCAGGATCAATCCCTTTGTGGACGCCAATCCCAATGCCAACAACTTCAATG 7655

y        967 --------------------ThrLeuValAlaGluThrCysIleSerIle 976
                                 ||||||          |||:::   |||
b       7656 GCAAAGGAGGTGTCTTTTAGACCCTATTTATATACTTGTGTGTGCATATA 7705
```

```
ESULT 8
S-09-221-013A-4
 Sequence 4, Application US/09221013A
 Patent No. 6495740
 GENERAL INFORMATION:
  APPLICANT:  Arioli, Antonio
  APPLICANT:  Williamson, Richard E.
  APPLICANT:  Betzner, Andreas S.
  APPLICANT:  Peng, Liangcai
  TITLE OF INVENTION:  Manipulation of cellulose and/or Beta-1,4-glucan
  FILE REFERENCE:  96-98
  CURRENT APPLICATION NUMBER: US/09/221,013A
  CURRENT FILING DATE:  1998-12-23
  PRIOR APPLICATION NUMBER:  PCT/AU97/00402
  PRIOR FILING DATE:  1997-06-24
  PRIOR APPLICATION NUMBER:  AU PO0699
  PRIOR FILING DATE:  1996-06-27
  NUMBER OF SEQ ID NOS:  37
  SOFTWARE:  PatentIn Ver. 2.0
, SEQ ID NO 4
,   LENGTH: 5009
,   TYPE: DNA
,   ORGANISM: Arabidopsis thaliana
,   FEATURE:
,   NAME/KEY: misc_feature
,   LOCATION: (1)..(5009)
,   OTHER INFORMATION: N is A,  T,  G  or C.
US-09-221-013A-4

Alignment Scores:
```

```
Pred. No.:              2.4e-[illegible]    Length:          5009
Score:                  231.[illegible]     Matches:         498
Percent Similarity:     52.[illegible]%     Conservative:    128
Best Local Similarity:  41.[illegible]%     Mismatches:      147
Query Match:            44.[illegible]%     Indels:          425
DB:                     4                   Gaps:            22

US-09-980-043-2 (1-978) x [illegible] (1-5009)

Qy          1 MetMetGluSerGlyAlaPr[illegible]IleCysHisThrCysGlyGluGlnValGlyHisAspAla 20
              :::  |||                |  :  !||||:::::::::
Db       1463 GTACAAGAACTGAG[illegible]AAACATTT[illegible]AAATCTGTGGAGATGAAATCGAATTAACGGTT 1522

Qy         21 AsnGlyGluLeuPr[illegible]ValAlaCysHisGluCysSerTyrProMetCysLysSerCysPhe 40
              :::  ||:|||| |||  ||| :::  :|||:::::::||:::||:::  :|||:::
Db       1523 AGCAGTGAGCTCT[illegible]AACAAATGCATTCCCGGTTTGTAGACCATGCTAT 1582

Qy         41 GluPheGlnIleA[illegible]rgSerArgArgArgCysArgCys---- ----          54
              |||:::\|  :|  :  |  |  :  :::||,
Db       1583 GAGTAC[illegible]AGTGCAAAACTCGATACAAAAGG 1642

Qy         55 ---- GlySerPr[illegible]TyrAspGluAsnLeuLeuAspAspVal 66
                  .|||:|.|,|                ::|||     :::|||||:::
Db       1643 ATTAAAG[illegible]AGTCCACGGGTTGATGGAGATGATGAAGAAGAAGAAGACATTGATGATCTT 1702

Qy         67 GluLysLys--  --  -  - GlySerGlyAsnGlnSerThrMetAlaSer 79
              |||  :::                          :::::  ::::::|||
Db       1703 GAGTATGA[illegible]TTTGATCATGGATGGACCTTGAACAGCCGCTGAAGCCGCACTCTCTTCA 1762

Qy         80 HisLeuAsnAspSerGlnAspValGlyIleHisAlaArg HisIleSer-  -  99
                |||  :::  :  :::  ::  |||::::::
Db       1763 CGCCTTAA[illegible]GTGGT  GGATTTGA[illegible]AGCTCCACCTGGCTCTCAGATTCCT 1819

Qy         95 --- ---  -  -  -  -------- -------  -  98
Db       1820 CTTTTGAGTTATT[illegible]AAGTGAGAATC[illegible]AAATTGTTTGTTTTCTCTGACAATGTTG 1879

Qy         96 -----  --  --  SerValSerThrValAspSerGluMetAsnAs 106
                          !|:::  |||::::::|,|
Db       1880 TTGCTTAGATGAT[illegible]TTTTTCTTATTAGTCTATGTGTTTTCAGGATGCTGATATGTATTC 1939

Qy        106 pGlu-  ----  ---  -  ---  - --------- ----  107
               :::
Db       1940 TGATCGTCATGCT[illegible]TATGTG[illegible]CTCTTCAAC[illegible]GGATATGGGAATCGCGTCTATCCTGC 1999

Qy        107 -----  ---  --  ---  -- --------- ---------- 107
Db       2000 ACCGTTTACAGATT[illegible]CTGCACCTCGTATGT[illegible]TTACTTTTATGATTCCTACAATTTTT 2059

Qy        107 -----  -----  --------  -- -----  ---------- 107
Db       2060 CTTCTTATATGATT[illegible]GTCAC[illegible]CTAATGAGT[illegible]ATGAAATGGTTTTGTTTGTTGTTTTC 2119

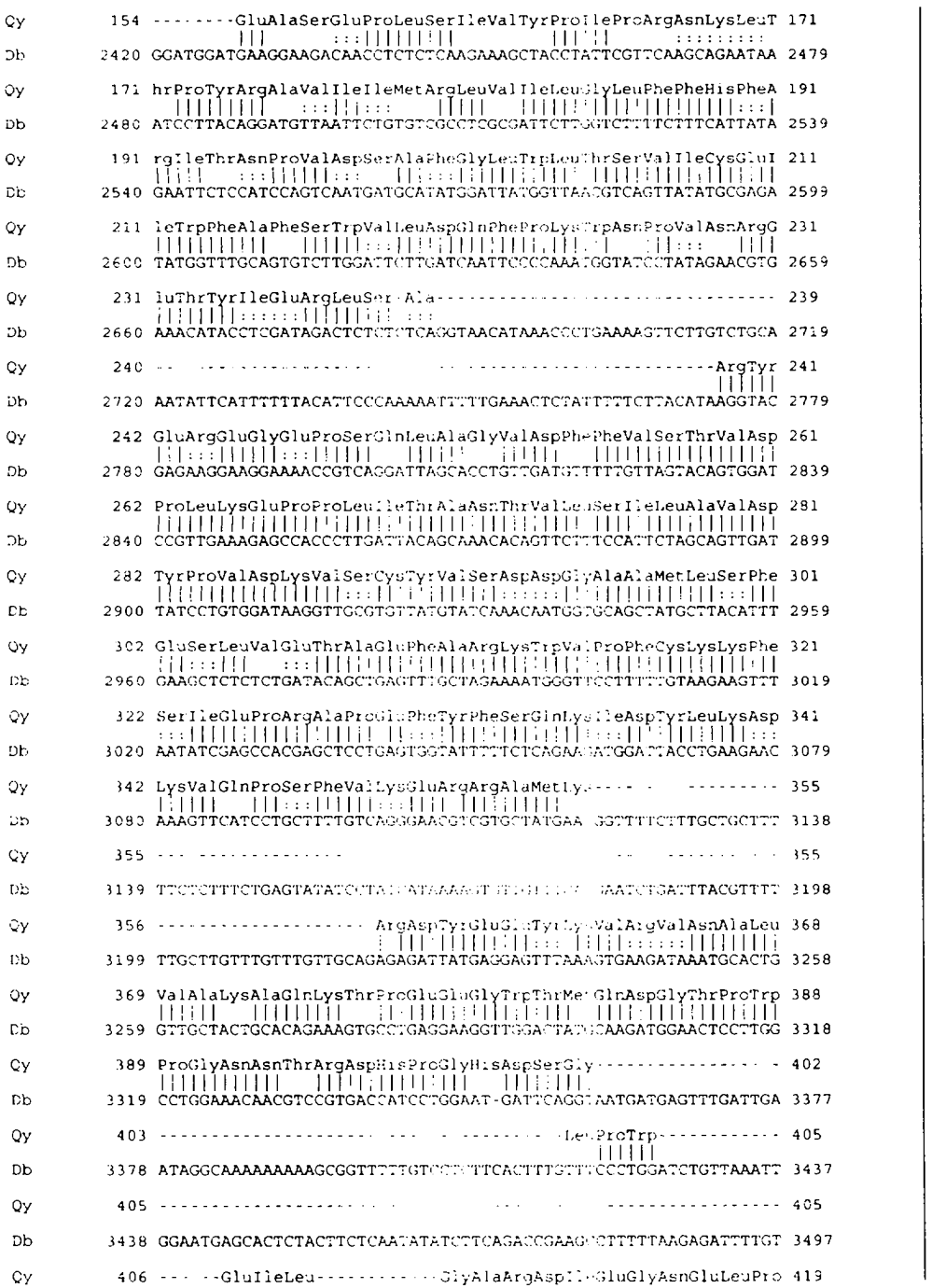
Qy        108 --------  -  --  -  ---- ----------TyrGlyAsnPr 111
                                                 |||||:::
Db       2120 AGCACAGGCGAGAT[illegible]AATGGTTCCTCAGAAAGATATTGCGGAATATGGTT[illegible]ATGGAAGTGT 2179

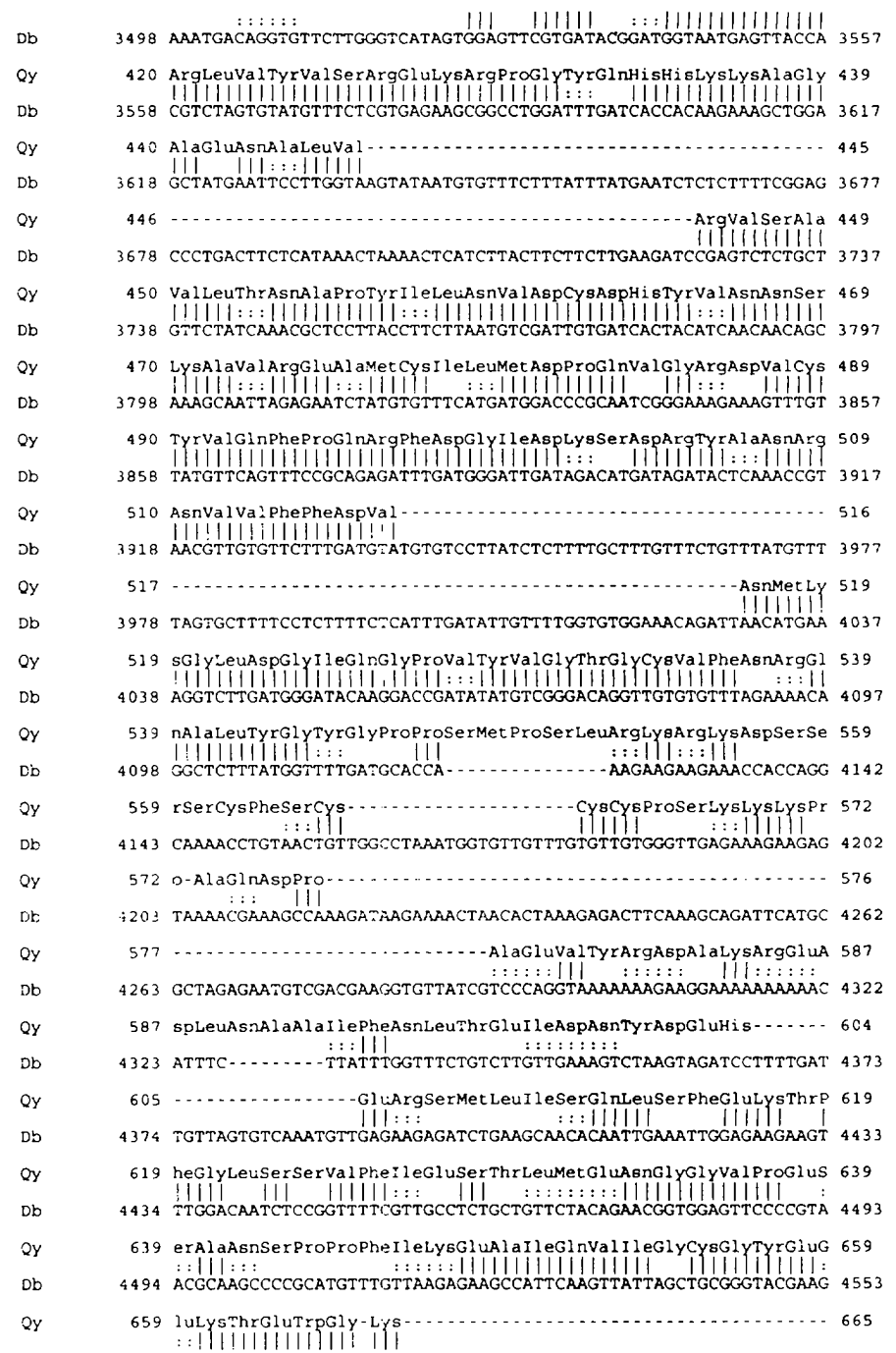
Qy        111 eIleTrpLysAsnArgVal LysSerCysLysAspLysGluAsnLysLysLysLysArgS 131
              :|||:::|, ::: :||  :::,||  :::  |||  |||
Db       2180 TGCTTGGAAGGAC[illegible]TATGGAAGTTTGGAAGAGACGACAAGGCGAAAAGCTTCAAGTCAT 2239

Qy        131 erProLysAlaGl[illegible]hrGluProAlaGlnValProThrGlu---------------  144
                  |||          :::|||||
Db       2240 TAAGCATGAAGGAG[illegible]AAACAATGGTCGAGGTTCCAATGATGACGACGAACTAGATGATCC 2299

Qy        145 ------------ GlnGlnMetGluGluLysProSerAla---------------  153
                          :::  :::::          |||:::
Db       2300 TGACATGCCTATGTAAGTTGTTAAAATCTAACAAAAGTTCAGATGAAATGATGCTCTGAA 2359

Qy        153 ------- -----  ---------  ------------------------------ -  153
Db       2360 ATTTTGTGTTCAATGGNTTTGTTTTCTTATTGTTGTTTAAACATTTTTCGTGCTAATTCA 2419
```





```
Db       4554 ATAAAACCGAATGGGGAAAAGAGGTAGAAAACATTACAAAGTTTTTCAACTTCTGAAAAC 4613

Qy        666 ----------------------------------------GlnIleGlyTrpIleTyr 671

Db       4614 TAGAAAAGTTCTTGTGATCTCATTCTTGCTGATAATCACACGCAGATCGGGTGGATTTAT 4673

Qy        672 GlySerValThrGluAspIleLeuSerGlyPheLysMetHisCysArgGlyTrpArgSer 691

Db       4674 GGATCGGTGACTGAAGATATCCTGACGGGTTTCAAGATGCATTGCCATGGATGGAGATCT 4733

Qy        692 IleTyrCysMetProValArgProAlaPheLysGlySerAlaProIleAsnLeuSerAsp 711

Db       4734 GTGTACTGTATGCCTAAGCGTGCAGCTTTTAAAGGATCTGCTCCTATTAACTTGTCAGAT 4793

Qy        712 ArgLeuHisGlnValLeuArgTrpAlaLeuGlySerValGluIlePhePheSerArgHis 731

Db       4794 CGTCTTCATCAAGTTCTACGTTGGGCTCTTGGCTCTGTAGAGATTTTCTTGAGCAGACAT 4853

Qy        732 CysProLeuTrpTyrGlyPheGlyGlyGlyArgLeuLysTrpLeuGlnArgLeuAlaTyr 751

Db       4854 TGTCCGATATGGTATGGTTATGGTGGTGGT---TTAAAATGGTTGGAGAGATTCTCTTAC 4910

Qy        752 IleAsnThrIleValTyrProPheThrSerLeuProLeuIleAlaTyrCysThrIlePro 771

Db       4911 ATCAACTCTGTCGTCTATCCTTGGACTTCACTTCCATTGATCGTCTATTGTTCTCTCCCC 4970

Qy        772 AlaValCysLeuLeuThrGlyLysPheIleIlePro 783

Db       4971 GCGGTTTGTTTACTCACAGGAAAATTCATCGTCCCT 5006
```

ESULT 9
S-09-221-013A-1
Sequence 1, Application US/09221013A
Patent No. 6495740
GENERAL INFORMATION:
APPLICANT: Arioli, Antonio
APPLICANT: Williamson, Richard E.
APPLICANT: Betzner, Andreas S.
APPLICANT: Peng, Liangcai
TITLE OF INVENTION: Manipulation of cellulose and/or Beta-1,4-glucan
FILE REFERENCE: 96-98
CURRENT APPLICATION NUMBER: US/09/221,013A
CURRENT FILING DATE: 1998-12-23
PRIOR APPLICATION NUMBER: PCT/AU97/00402
PRIOR FILING DATE: 1997-06-24
PRIOR APPLICATION NUMBER: AU PO0699
PRIOR FILING DATE: 1996-06-27
NUMBER OF SEQ ID NOS: 37
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 2248
TYPE: DNA
ORGANISM: Arabidopsis thaliana
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(1887)
S-09-221-013A-1

```
Alignment Scores:
Pred. No.:                5.75e-248      Length:          2248
Score:                    2284.50        Matches:         430
Percent Similarity:       77.18%         Conservative:    74
Best Local Similarity:    65.85%         Mismatches:      108
Query Match:              43.60%         Indels:          41
DB:                       4              Gaps:            10

US-09-980-043-2 (1-978) x US-09-221-013A-1 (1-2248)

Qy        352 ArgAlaMetLysArgAspTyrGluGluTyrLysValArgValAsnAlaLeuValAlaLys 371

Db          1 CGAGCTATGAAGAGAGAGTATGAAGAGTTTAAAGTGAGGATAAATGCTCTTGTTGCCAAA 60

Qy        372 AlaGlnLysThrProGluGluGlyTrpThrMetGlnAspGlyThrProTrpProGlyAsn 391

Db         61 GCACAGAAAATCC[illegible]GGAGAAGGCTCAACAATGCAGGATGGTACTCCCTGGCCTGGTAAC 120

Qy        392 AsnThrArgAspHisPro[illegible]GlyHisAspSerGlyLeuPro 404

Db        121 AACACTAGAGATCA[illegible]GAATGATACA[illegible]TTCTTAGGCCATAGTGGGGGTCTG - 177

Qy        405 TrpGluIleLeuG[illegible]AlaArgAsp[illegible]GlyAsnGluLeuProArgLeuValTyrVal 424

Db        178 --- ----  [illegible]GATA[illegible]GATGGAAATGAGCTGCCTAGACTCATCTATGTT 216

Qy        425 SerArgGluLysAr[illegible]GlyTyrGlnHisHisLysLysAlaGlyAlaGluAsnAlaLeu 444

Db        217 TCTCGTGAAAAGC[illegible]CTTGATTTCAACA[illegible]CACAAAAAGGCTGGAGCTATGAATGCATCG 276

Qy        445 ValArgValSerAl[illegible]AsnAl[illegible]ProTyrIleLeuAsnValAspCysAspHis 464

Db        277 ATCCGT[illegible]ATC[illegible]  [illegible]ATCTTTTGAAGGTGGATTGTGATCAT 336

Qy        465 TyrValAsnAsn[illegible]AlaMet[illegible]CysIleLeuMetAspProGlnVal 484

Db        337 TACTTTAATAACA[illegible]AAGCTATGAAAGAAGCTATGTTTCATGATGGACCCGGCTATT 396

Qy        485 GlyArgAspValCysTyrValGlnPhePro[illegible]GlnArgPheAspGlyIleAspLysSerAsp 504

Db        397 GGAAAGAAGTGCTG[illegible]TATGTCCA[illegible]CCTCAACGTTTTGACGGTATTGATTTGCACGAT 456

Qy        505 ArgTyrAlaAsnAr[illegible]AsnValValPhePheAspValAsnMetLysGlyLeuAspGlyIle 524

Db        457 CGATATGC[illegible]AACA[illegible]AATATAGTCTTTTTCGATATTAACATGAAGGGGTTGGATGGTATC 516

Qy        525 GlnGlyProValTyrValGlyThrGlyCysVal[illegible]PheAsnArgGlnAlaLeuTyrGlyTyr 544

Db        517 CACGGTCCAGTATATG[illegible]TGGGTACTGGT[illegible]TTTTAATAGGCAGGCTCTATATGGGTAT 576

Qy        545 GlyProProSerMet[illegible]ProSerLeuArgLysArgLysAspSerSerSerCysPheSerCys 564

Db        577 GATCCT- --- - GTTTTGACGGAAGAAGATTTAGAACCAAATATTATTGTCAAGAGC 627

Qy        565 CysCysProSerLysLysLysProAlaGlnAspProAlaGluValTyr---------Arg 581

Db        628 TGTTGCGGGTCAAG[illegible]AAGAAAGGTAAAA[illegible]CAAGAAGTATAACTACGAAAAGAGGAGA 687

Qy        582 AspAlaLysArgGl[illegible]AspLeuAsnAlaA[illegible]IlePheAsnLeuThrGluIleAsp------ 599

Db        688 GGCATCAACAGAAG[illegible]GACTCCAATGCTCCAC[illegible]TTTCAATATGGAGGACATCGATGAGGGT 747

Qy        600 ------AsnTyrAspGluHisGluArgSerMetLeuIleSerGlnLeuSerPheGluLys 617

Db        748 TTTGAAGGTTATGATGAT---GAGAGGTCTATTCTAATGTCCCAGAGGAGTGTAGAGAAG 804

Qy        618 ThrPheGlyLeuSer[illegible]SerValPheIleGluSer[illegible]ThrLeuMetGluAsnGlyGlyValPro 637

Db        805 CGT[illegible]GGT[illegible]CAGTC[illegible]CGGTATTTATTG[illegible]GCAACCTTCATGGAACAAGGCGGCATTCCA 864

Qy        638 GluSerAlaAsnSer[illegible]ProPhe[illegible]LysGluAlaIleGlnValIleGlyCysGlyTyr 657

Db        865 CCAACAACCAATCC[illegible]CTACTCTTCTGAAG[illegible]TATTCATGTTATAAGCTGTGGTTAC 924

Qy        658 GluGluLysThrGl[illegible]TrpGlyLysGlnIleGlyTrpIleTyrGlySerValThrGluAsp 677

Db        925 GAAGACAAGACTGAATGGGG[illegible]AAAGAGATTGGTT[illegible]GATCTATGGTTCCGTGACGGAAGAT 984

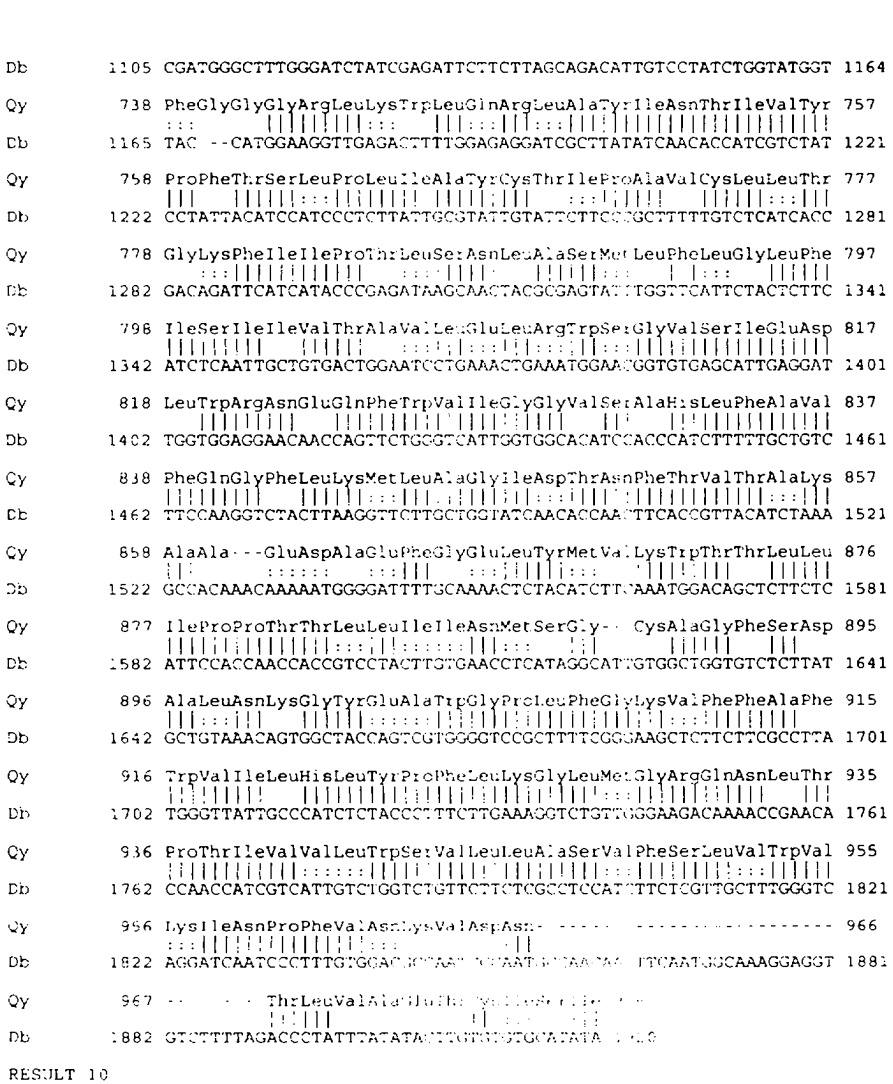
Qy        678 IleLeuSerGlyPheLysMetHisCysArgGlyTrpArgSerIleTyrCysMetProVal 697

Db        985 ATTCTTACTGGGTTCAAGATGCATGCCCGGGGTTGGATATCGATCTACTGCAATCCTCCA 1044

Qy        698 ArgProAlaPheLysGlySerAlaProIleAsnLeuSerAspArgLeuHisGlnValLeu 717

Db       1045 CGCCCTGCGTTCAAGGGATCTGCACCAATCAATCTTTCTGATCGTTTGAACCAAGTTCTT 1104

Qy        718 ArgTrpAlaLeuGlySerValGluIlePhePheSerArgHisCysProLeuTrpTyrGly 737
```

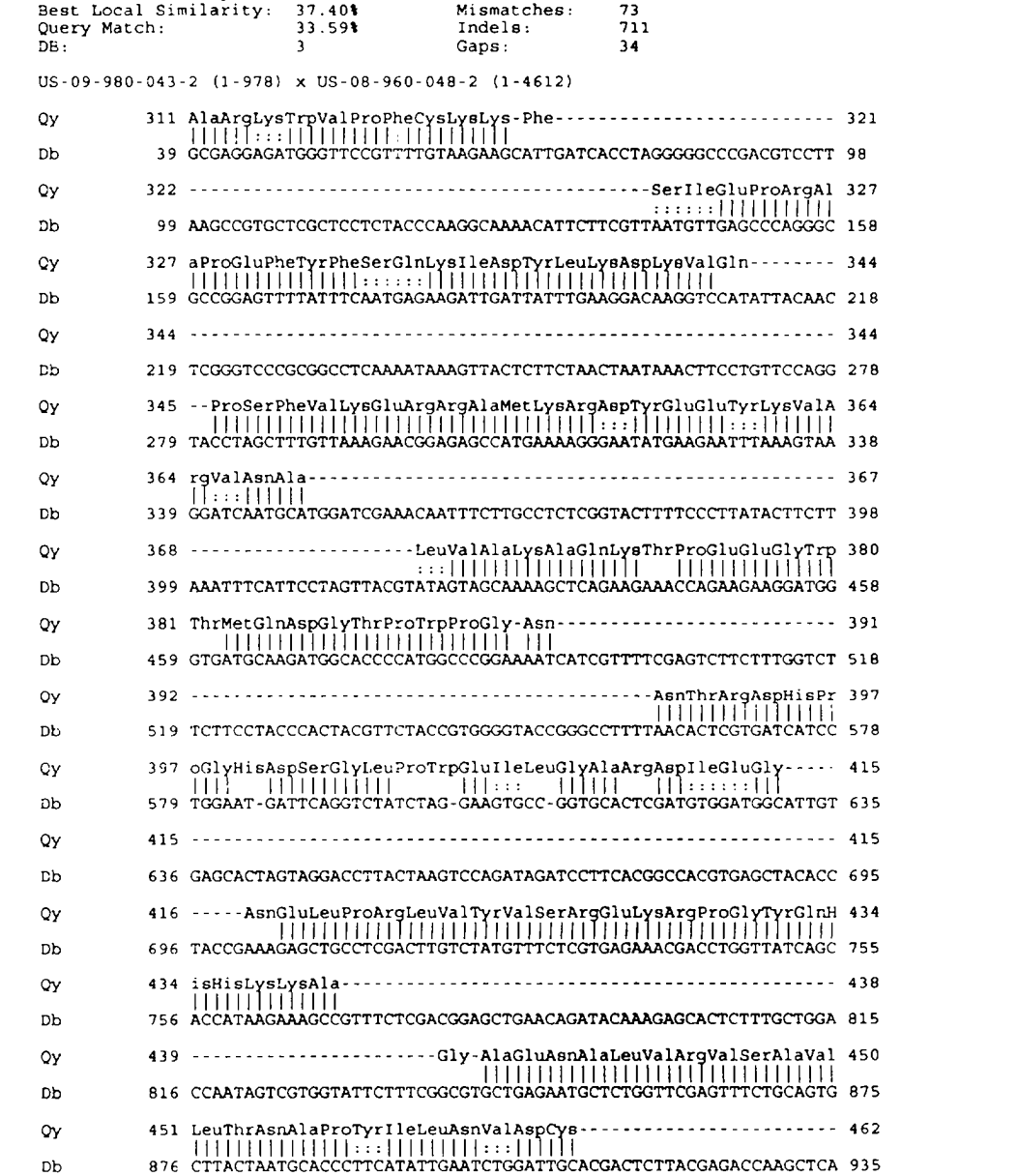


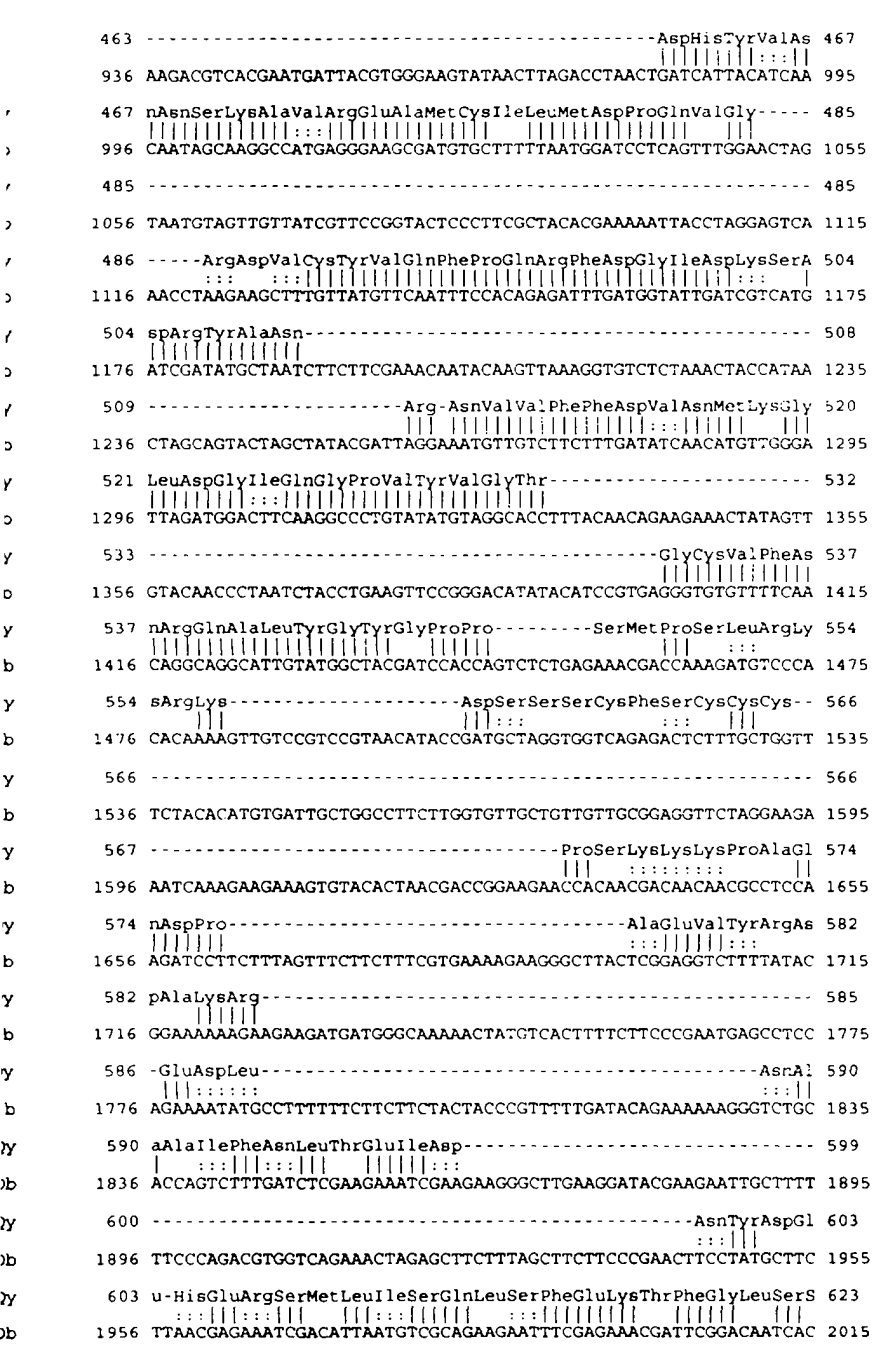
RESULT 10
US-08-960-048-2
; Sequence 2, Application US/08960048C
; Patent No. 6271443
; GENERAL INFORMATION:
; APPLICANT: Stalker, D. et al.
; TITLE OF INVENTION: Plant Cellulose Synthase and Promoter
; TITLE OF INVENTION: Sequences
; FILE REFERENCE: 15621/01/US
; CURRENT APPLICATION NUMBER: US/08/960,048C
; CURRENT FILING DATE: 1997-10-29
; PRIOR APPLICATION NUMBER: 60/029,987
; PRIOR FILING DATE: 1996-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 4612
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Oligonucleotide
US-08-960-048-2

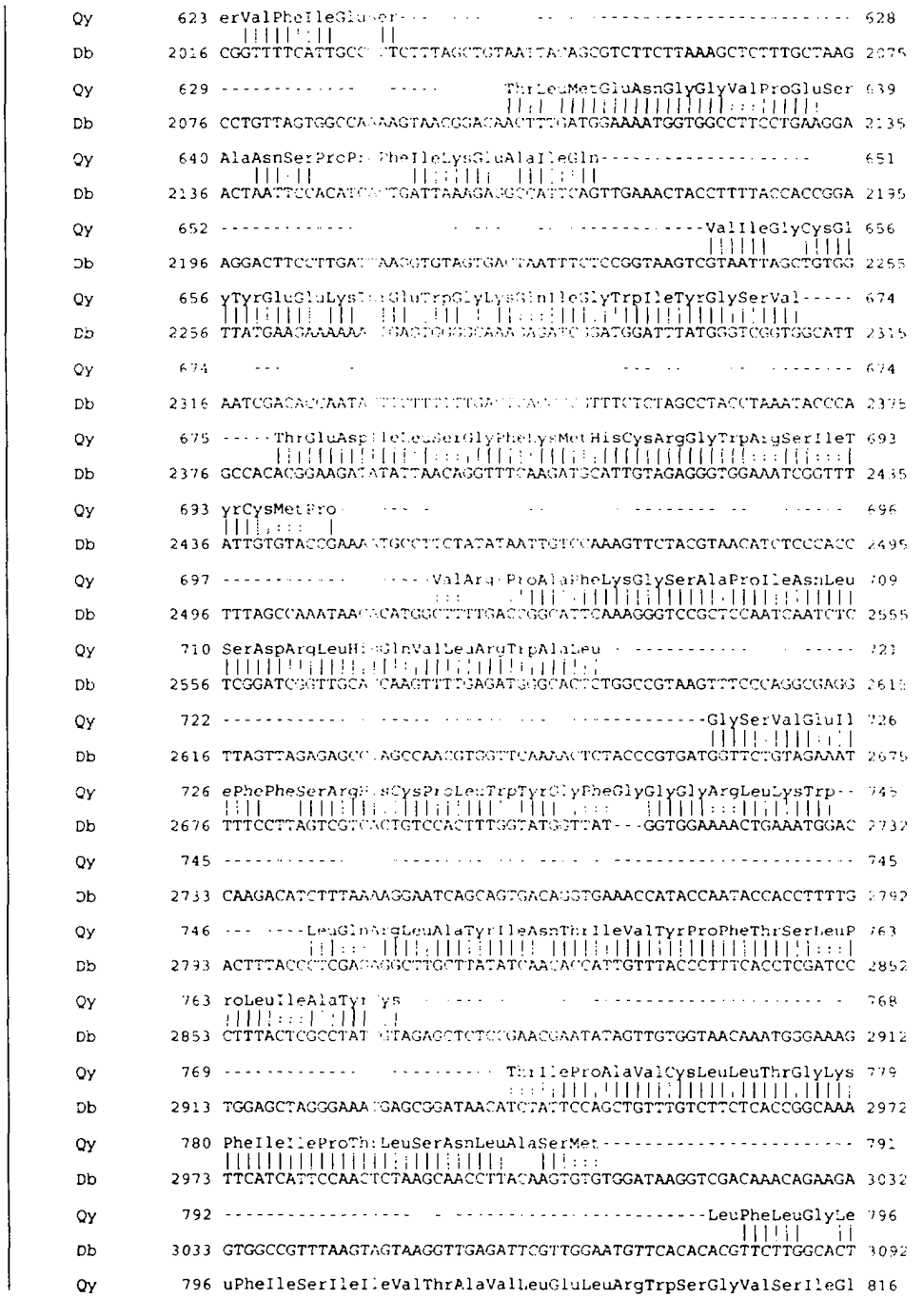
Alignment Scores:

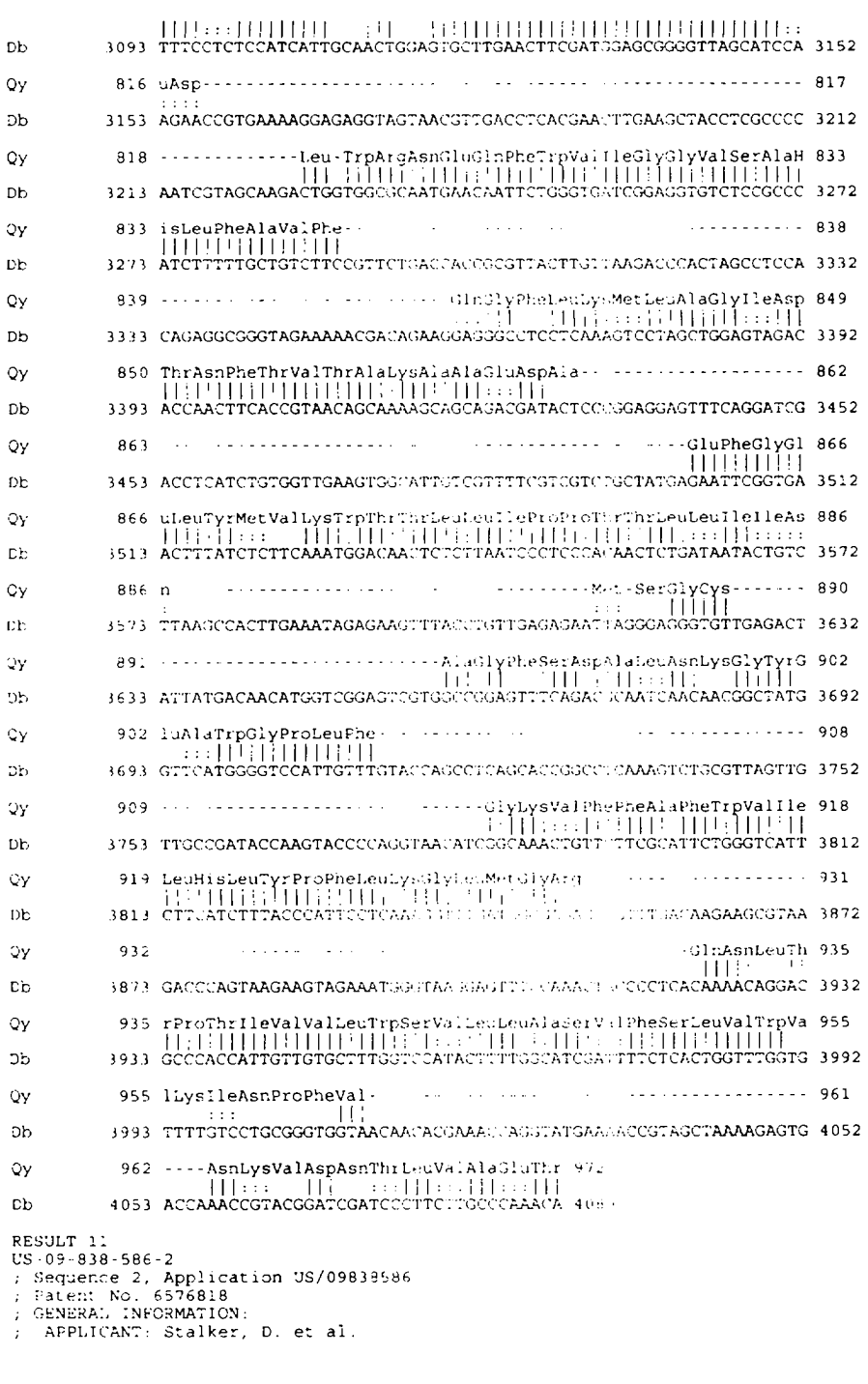
Pred. No.:	5.47e-188	Length:	4612
Score:	1760.00	Matches:	512
Percent Similarity:	42.95%	Conservative:	76
Best Local Similarity:	37.40%	Mismatches:	73
Query Match:	33.59%	Indels:	711
DB:	3	Gaps:	34

US-09-980-043-2 (1-978) x US-08-960-048-2 (1-4612)




463 ----------AspHisTyrValAs 467
936 AAGACGTCACGAATGATTACGTGGGAAGTATAACTTAGACCTAACTGATCATTACATCAA 995
467 nAsnSerLysAlaValArgGluAlaMetCysIleLeuMetAspProGlnValGly----- 485
996 CAATAGCAAGGCCATGAGGGAAGCGATGTGCTTTTTAATGGATCCTCAGTTTGGAACTAG 1055
485 ---------- 485
1056 TAATGTAGTTGTTATCGTTCCGGTACTCCCTTCGCTACACGAAAAATTACCTAGGAGTCA 1115
486 -----ArgAspValCysTyrValGlnPheProGlnArgPheAspGlyIleAspLysSerA 504
1116 AACCTAAGAAGCTTTGTTATGTTCAATTTCCACAGAGATTTGATGGTATTGATCGTCATG 1175
504 spArgTyrAlaAsn---------- 508
1176 ATCGATATGCTAATCTTCTTCGAAACAATACAAGTTAAAGGTGTCTCTAAACTACCATAA 1235
509 ----------Arg-AsnValValPhePheAspValAsnMetLysGly 520
1236 CTAGCAGTACTAGCTATACGATTAGGAAATGTTGTCTTCTTTGATATCAACATGTTGGGA 1295
521 LeuAspGlyIleGlnGlyProValTyrValGlyThr---------- 532
1296 TTAGATGGACTTCAAGGCCCTGTATATGTAGGCACCTTTACAACAGAAGAAACTATAGTT 1355
533 ----------GlyCysValPheAs 537
1356 GTACAACCCTAATCTACCTGAAGTTCCGGGACATATACATCCGTGAGGGTGTGTTTTCAA 1415
537 nArgGlnAlaLeuTyrGlyTyrGlyProPro---------SerMetProSerLeuArgLy 554
1416 CAGGCAGGCATTGTATGGCTACGATCCACCAGTCTCTGAGAAACGACCAAAGATGTCCCA 1475
554 sArgLys----------AspSerSerSerCysPheSerCysCysCys-- 566
1476 CACAAAAGTTGTCCGTCCGTAACATACCGATGCTAGGTGGTCAGAGACTCTTTGCTGGTT 1535
566 ---------- 566
1536 TCTACACATGTGATTGCTGGCCTTCTTGGTGTTGCTGTTGTTGCGGAGGTTCTAGGAAGA 1595
567 ----------ProSerLysLysLysProAlaGl 574
1596 AATCAAAGAAGAAAGTGTACACTAACGACCGGAAGAACCACAACGACAACAACGCCTCCA 1655
574 nAspPro----------AlaGluValTyrArgAs 582
1656 AGATCCTTCTTTAGTTTCTTCTTTCGTGAAAAGAAGGGCTTACTCGGAGGTCTTTTATAC 1715
582 pAlaLysArg---------- 585
1716 GGAAAAAAGAAGAAGATGATGGGCAAAAACTATGTCACTTTTCTTCCCGAATGAGCCTCC 1775
586 -GluAspLeu----------AsnAl 590
1776 AGAAAATATGCCTTTTTTCTTCTTCTACTACCCGTTTTTGATACAGAAAAAAGGGTCTGC 1835
590 aAlaIlePheAsnLeuThrGluIleAsp---------- 599
1836 ACCAGTCTTTGATCTCGAAGAAATCGAAGAAGGGCTTGAAGGATACGAAGAATTGCTTTT 1895
600 ----------AsnTyrAspGl 603
1896 TTCCCAGACGTGGTCAGAAACTAGAGCTTCTTTAGCTTCTTCCCGAACTTCCTATGCTTC 1955
603 u-HisGluArgSerMetLeuIleSerGlnLeuSerPheGluLysThrPheGlyLeuSerS 623
1956 TTAACGAGAAATCGACATTAATGTCGCAGAAGAATTTCGAGAAACGATTCGGACAATCAC 2015


Qy 623 erValPheIleGlu---------- 628
Db 2016 CGGTTTTCATTGCC TCTTTAGCTGTAATTAAAGCGTCTTCTTAAAGCTCTTTGCTAAG 2075
Qy 629 ---------- ThrLeuMetGluAsnGlyValProGluSer 639
Db 2076 CCTGTTAGTGGCCA AAGTAACGGACAACTTTGATGGAAAATGGTGGCCTTCCTGAAGGA 2135
Qy 640 AlaAsnSerProP PheIleLysGluAlaIleGln---------- 651
Db 2136 ACTAATTCCACATC TGATTAAAGAGCCATTCAGTTGAAACTACCTTTTACCACCGGA 2195
Qy 652 ---------- ----------ValIleGlyCysGl 656
Db 2196 AGGACTTCCTTGAT AAGGTGTAGTGA TAATTTCTCCGGTAAGTCGTAATTAGCTGTGG 2255
Qy 656 yTyrGluGluLys GluTrpGlyLysGlnIleGlyTrpIleTyrGlySerVal----- 674
Db 2256 TTATGAAGAAAAAA GAGTGGGGAAA AGATGGATTTATGGGTCGGTGGCATT 2315
Qy 674 ---------- 674
Db 2316 AATCGACAGCAATA TTTTTTTTGA TTTTCTCTAGCCTACCTAAATACCCA 2375
Qy 675 -----ThrGluAsp IleLeuSerGlyPheLysMetHisCysArgGlyTrpArgSerIleT 693
Db 2376 GCCACACGGAAGAT ATATTAACAGGTTTCAAGATGCATTGTAGAGGGTGGAAATCGGTTT 2435
Qy 693 yrCysMetPro---------- 696
Db 2436 ATTGTGTACCGAAA TGCCTTCTATATAATTGTCCAAAGTTCTACGTAACATCTCCCACC 2495
Qy 697 ---------- -----ValArg ProAlaPheLysGlySerAlaProIleAsnLeu 709
Db 2496 TTTAGCCAAATAAC ATGGCTTTTGACGGCATTCAAAGGGTCCGCTCCAATCAATCTC 2555
Qy 710 SerAspArgLeuH GlnValLeuArgTrpAlaLeu---------- ----- 721
Db 2556 TCGGATCGGTTGCA CAAGTTTGAGATGGCACTCTGGCCGTAAGTTTCCCAGGCGAGG 2615
Qy 722 ---------- ----------GlySerValGluIl 726
Db 2616 TTAGTTAGAGAGCC AGCCAACGTGGTTCAAAACTCTACCCGTGATGGTTCTGTAGAAAT 2675
Qy 726 ePhePheSerArg sCysProLeuTrpTyrGlyPheGlyGlyArgLeuLysTrp-- 745
Db 2676 TTTCCTTAGTCGTCACTGTCCACTTTGGTATGGTTAT---GGTGGAAAACTGAAATGGAC 2732
Qy 745 ---------- 745
Db 2733 CAAGACATCTTTAAAAGGAATCAGCAGTGACAGGTGAAACCATACCAATACCACCTTTTG 2792
Qy 746 --- ----LeuGlnArgLeuAlaTyrIleAsnThrIleValTyrProPheThrSerLeuP 763
Db 2793 ACTTTACCCTCGA AGGCTTGCTTATATCAACACCATTGTTTACCCTTTCACCTCGATCC 2852
Qy 763 roLeuIleAlaTyr ys---------- 768
Db 2853 CTTTACTCGCCTAT TAGAGCTCTCCGAACGAATATAGTTGTGGTAACAAATGGGAAAG 2912
Qy 769 ---------- ----------ThrIleProAlaValCysLeuLeuThrGlyLys 779
Db 2913 TGGAGCTAGGGAAA GAGCGGATAACATCTATTCCAGCTGTTTGTCTTCTCACCGGCAAA 2972
Qy 780 PheIleIleProTh LeuSerAsnLeuAlaSerMet---------- 791
Db 2973 TTCATCATTCCAACTCTAAGCAACCTTACAAGTGTGTGGATAAGGTCGACAAACAGAAGA 3032
Qy 792 ---------- ----------LeuPheLeuGlyLe 796
Db 3033 GTGGCCGTTTAAGTAGTAAGGTTGAGATTCGTTGGAATGTTCACACACGTTCTTGGCACT 3092
Qy 796 uPheIleSerIleIleValThrAlaValLeuGluLeuArgTrpSerGlyValSerIleGl 816

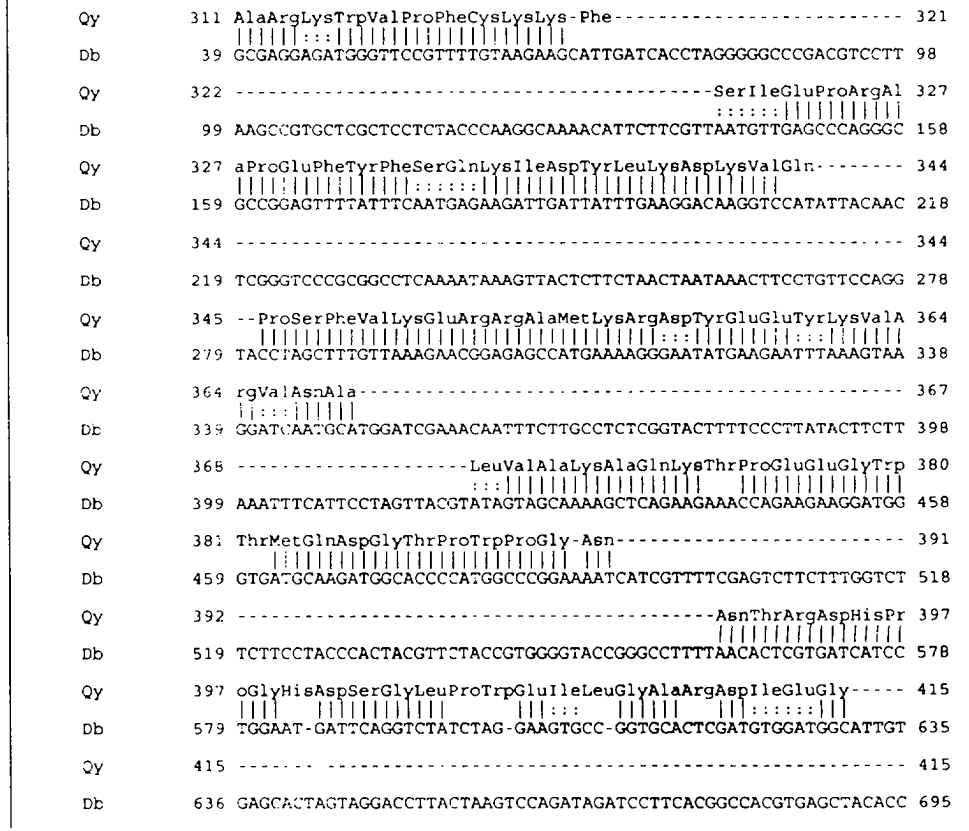


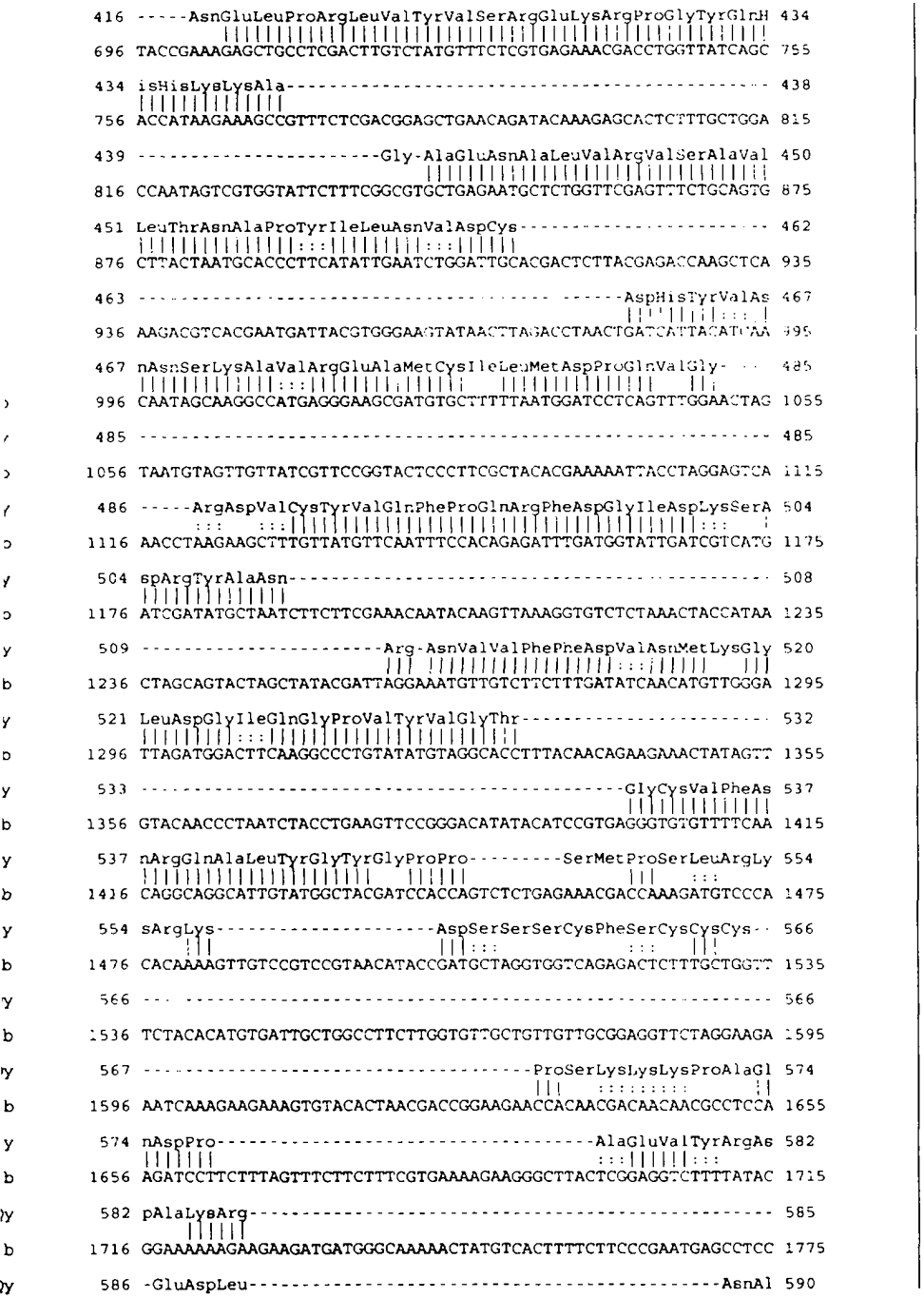
RESULT 11
US-09-838-586-2
; Sequence 2, Application US/09838586
; Patent No. 6576818
; GENERAL INFORMATION:
; APPLICANT: Stalker, D. et al.
; TITLE OF INVENTION: Plant Cellulose Synthase and Promoter
; TITLE OF INVENTION: Sequences
; FILE REFERENCE: 15621/02/US
; CURRENT APPLICATION NUMBER: US/09/838,586
; CURRENT FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/029,987
; PRIOR FILING DATE: 1996-10-29
; PRIOR APPLICATION NUMBER: 08/960,048
; PRIOR FILING DATE: 1997-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 4612
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Oligonucleotide
US-09-838-586-2

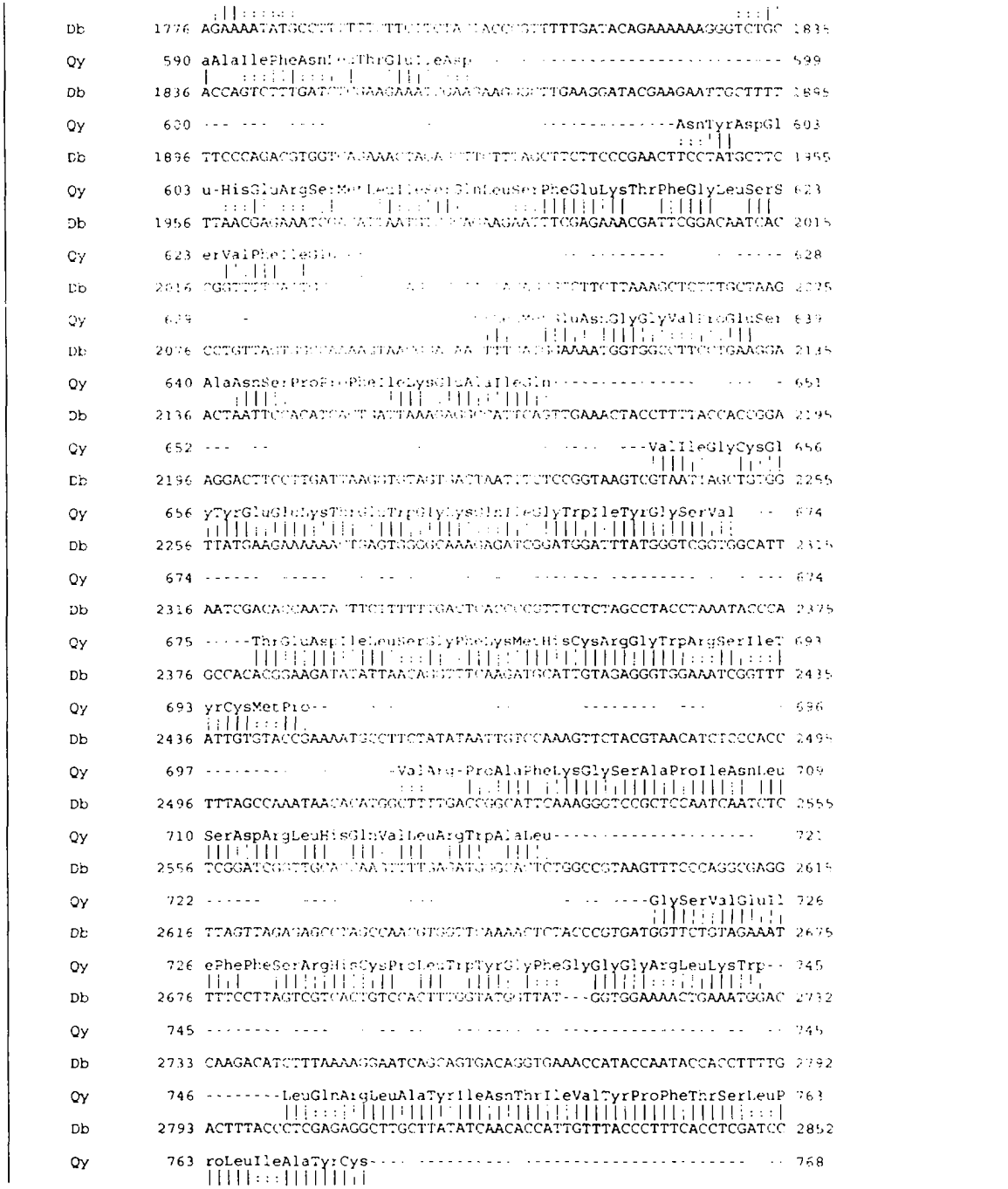
Alignment Scores:

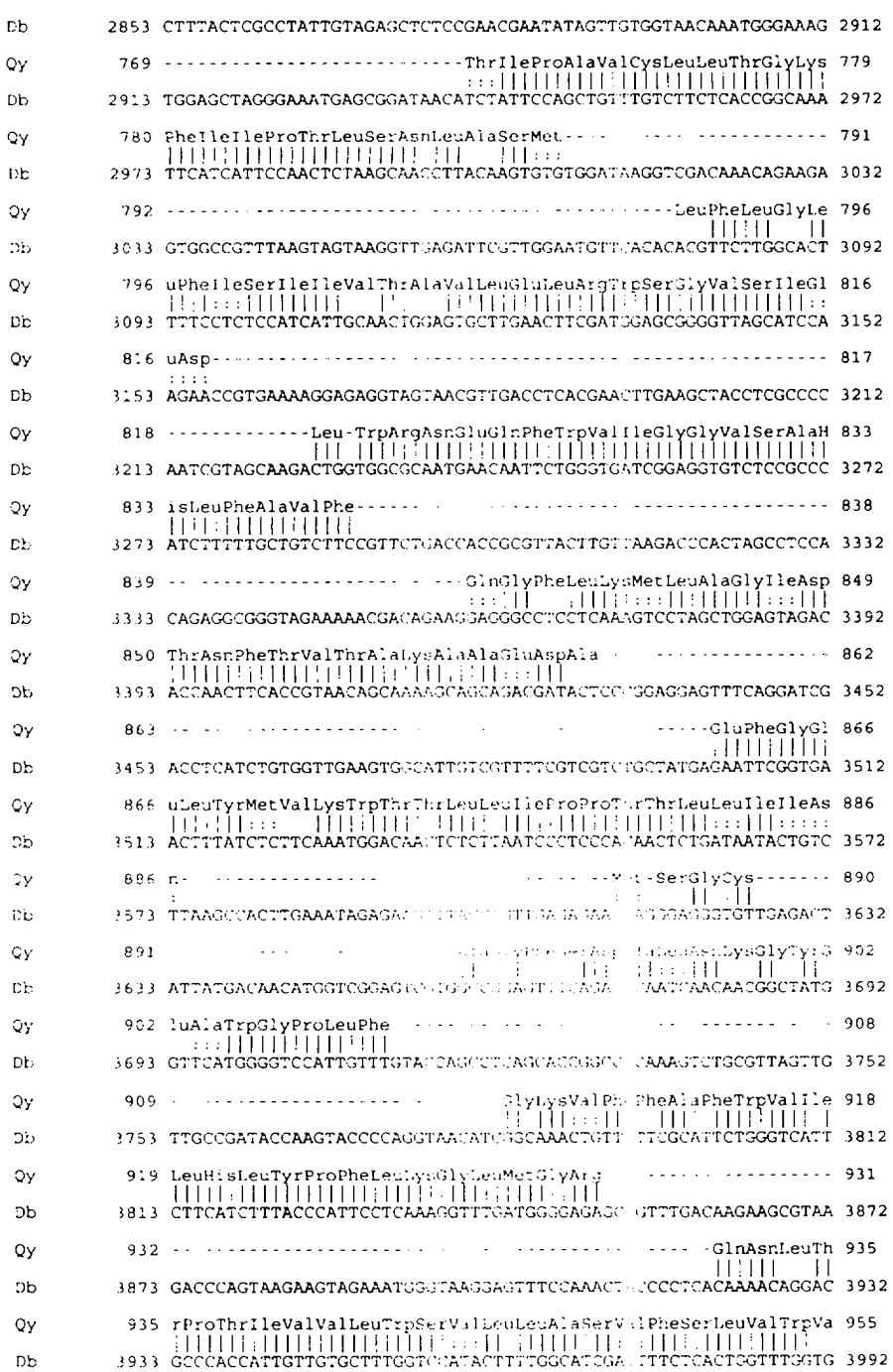
Pred. No.:	5.47e-188	Length:	4612
Score:	1760.00	Matches:	512
Percent Similarity:	42.95%	Conservative:	76
Best Local Similarity:	37.40%	Mismatches:	73
Query Match:	33.59%	Indels:	711
DB:	4	Gaps:	34

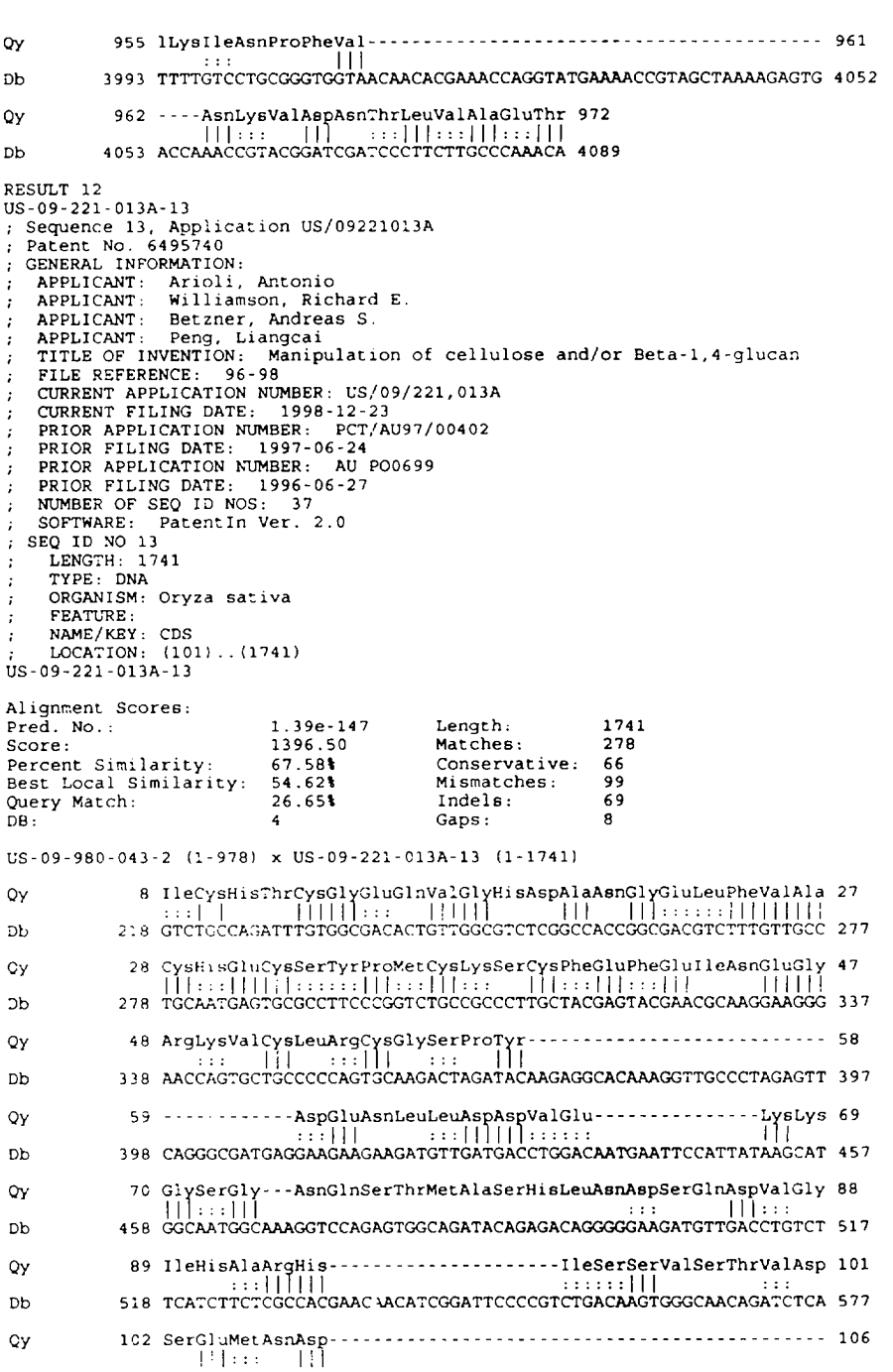
US-09-980-043-2 (1-978) x US-09-838-586-2 (1-4612)











RESULT 12
US-09-221-013A-13
; Sequence 13, Application US/09221013A
; Patent No. 6495740
; GENERAL INFORMATION:
; APPLICANT: Arioli, Antonio
; APPLICANT: Williamson, Richard E.
; APPLICANT: Betzner, Andreas S.
; APPLICANT: Peng, Liangcai
; TITLE OF INVENTION: Manipulation of cellulose and/or Beta-1,4-glucan
; FILE REFERENCE: 96-98
; CURRENT APPLICATION NUMBER: US/09/221,013A
; CURRENT FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: PCT/AU97/00402
; PRIOR FILING DATE: 1997-06-24
; PRIOR APPLICATION NUMBER: AU PO0699
; PRIOR FILING DATE: 1996-06-27
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 1741
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (101)..(1741)
US-09-221-013A-13

Alignment Scores:

Pred. No.:	1.39e-147	Length:	1741
Score:	1396.50	Matches:	278
Percent Similarity:	67.58%	Conservative:	66
Best Local Similarity:	54.62%	Mismatches:	99
Query Match:	26.65%	Indels:	69
DB:	4	Gaps:	8

US-09-980-043-2 (1-978) x US-09-221-013A-13 (1-1741)

```
      578 GGAGAGATCCCTGATGCTTCCCCCGATCGCCATTCTATCCGCAGCGGAACATCAAGCTAT 637

      107 ------------------------------------------------------------Glu 107

      638 GTTGATCCAAGTGTTCCAGTTCCTGTGAGGATTGTGGACCCCTCCAAGGACTTGAATTCC 697

      108 TyrGly---AsnProIle---TrpLysAsnArgValLysSerCysLysAspLysGluAsn 125
          |||||||   |||   :::   |||:::   |||||||   |||   :::::::|||:::::::
      698 TATGGGATTAACAGTGTTGACTGGCAAGAAAGAGTTGCCAGCTGGAGGAACAAGCAGGAC 757

      126 LysLys------------LysLysArgSerProLysAlaGluThrGluProAlaGlnVal 141
          |||                 :::   |||:::|||                  :::
      758 AAAAATATGATGCAGGTAGCTAATAAATATCCAGAGGCAAGAGGGGGAGACATGGAAGGG 817

      142 ProThrGluGlnGlnMetGluGluLysProSerAlaGluAlaSerGluProLeuSerIle 161
                      :::   :::          :::|||      :||||||||
      818 ACTGGTTCAAATGGTGAAGATATCCAAATGGTTGATGATGCACGTCTACCTCTGAGCCGC 877

      162 ValTyrProIleProArgAsnLysLeuThrProTyrArgAlaValIleIleMetArgLeu 181
          :::   |||||||||   |||:::|||       ||||||   ||||||:||:::||||||
      878 ATAGTGCCTATCCCTTCAAACCAGCTCAACCTTTACCGGATTGTTATCATTCTCCGTCTT 937

      182 ValIleLeuGlyLeuPhePheHisPheArgIleThrAsnProValAspSerAlaPheGly 201
          :::||||||       ||||||   :::|||:::|||:::||||||       |||:::|||
      938 ATCATCCTGATGTTCTTCTTCCAATATCGTGTCACTCATCCAGTGCGGGATGCTTATGGA 997

      202 LeuTrpLeuThrSerValIleCysGluIleTrpPheAlaPheSerTrpValLeuAspGln 221
          |||||||||   |||||||||||||||||||||          ||||||:::|||||||||
      998 TTGTGGCTAGTATCTGTTATCTGTGAAATTTGGTTGCCCTTATCCTGGCTCCTAGATCAA 1057

      222 PheProLysTrpAsnProValAsnArgGluThrTyrIleGluArgLeuSerAlaArgTyr 241
          |||||||||||   |||:::||||||||||||||||:::::||||||:::   ||||||
     1058 TTCCCAAAGTGGTACCCGATAAACCGTGAAACATACCTTGACAGGCTTGCATTGAGATAT 1117

      242 GluArgGluGlyGluProSerGlnLeuAlaGlyValAspPhePheValSerThrValAsp 261
          :::|||||||||||||||||||||||||||||   :::|||   |||||||||||||||||
     1118 GATAGGGAGGGAGAGCCATCACAGCTTGCTCCCATTGATGTCTTTGTCAGTACGGTGGAT 1177

      262 ProLeuLysGluProProLeuIleThrAlaAsnThrValLeuSerIleLeuAlaValAsp 281
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
     1178 CCACTAAAGGAACCTCCTCTGATCACAGCAAACACTGTTTTGTCCATTCTGGCTGTGGAT 1237

      282 TyrProValAspLysValSerCysTyrValSerAspAspGlyAlaAlaMetLeuSerPhe 301
          |||||||||||||||||||||||||||||||||||||:::|||||||:::|||||||:::|||
     1238 TACCCTGTTGACAAAGTGTCATGCTATGTTTCTGACGATGGTTCAGCTATGTTAACTTTT 1297

      302 GluSerLeuValGluThrAlaGluPheAlaArgLysTrpValProPheCysLysLysPhe 321
          |||:::|||   ||||||||||||||||||||||||||||||||||||||||||||||||
     1298 GAGGCTCTGTCAGAAACTGCAGAATTTGCTAGGAAGTGGGTTCCGTTTTGCAAGAAGCAC 1357

      322 SerIleGluProArgAlaProGluPheTyrPheSerGlnLysIleAspTyrLeuLysAsp 341
          :::|||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||
     1358 AATATTGAACCACGAGCTCCAGAGTTTTACTTTGCTCAAAAAATAGATTACCTGAAGGAC 1417

      342 LysValGlnProSerPheValLysGluArgArgAlaMetLysArgAspTyrGluGluTyr 361
          |||:::|||||||||||||||||||||||||||||||||||||||||||||||||||:::
     1418 AAAATCCAACCTTCCTTTGTTAAAGAAAGGCGGGCAATGAAGAGAGAGTATGAAGAATTC 1477

      362 LysValArgValAsnAlaLeuValAlaLysAlaGlnLysThrProGluGluGlyTrpThr 381
          |||||||||:::|||||||||||||||||||||||||||   |||||||||||||||||||
     1478 AAGGTACGGATCAATGCTCTTGTTGCGAAGGCACAAAAAGTACCTGAAGAGGGGTGGACC 1537

      382 MetGlnAspGlyThrProTrpProGlyAsnAsnThrArgAspHisProGlyHisAspSer 401
          |||   |||||||||   ||||||||||||||||   ||||||||||||   ||||||||||
     1538 ATGGCTGATGGCACTGCTTGGCCTGGGAATAACCCAAGGGATCACCCTGG-CATGATTCA 1596

      402 GlyLeuProTrpGluIleLeuGlyAlaArgAspIleGluGlyAsnGluLeuProArgLeu 421
          |||:::   |||         |||       |||   :::|||||||||||||||||||||
     1597 GGTGTTCT-TGGGGCACAGT-GGTGGGCTTGACACTGATGGTAACGAGTTGCCACGGCTT 1654

      422 ValTyrValSerArgGluLysArgProGlyTyrGlnHisHisLysLysAlaGlyAlaGlu 441
          ||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||||
     1655 GTCTACGTCTCTCGTGAAAAGAGGCCAGGATTCCAGCATCACAAGAAGGCTGGTGCAATG 1714
```

```
Qy       442 AsnAlaLeuValArgValSerAlaVal 469
             |||:|||:.:::|| ||||||||:|||
Db      1715 AATGCATTGATTCGTGTATCTGCTGTC 1741

RESULT 13
US-09-313-294A-5519
; Sequence 5519, Application: US/09313294A
; Patent No. 6476212
; GENERAL INFORMATION:
;  APPLICANT:  Lalgudi, Raghunath V
;  APPLICANT:  Ito, Laura Y.
;  APPLICANT:  Sherman, Bradley K.
;  TITLE OF INVENTION:  POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
;  FILE REFERENCE:  PL-0017 US
;  CURRENT APPLICATION NUMBER:  US/09/313,294A
;  CURRENT FILING DATE:  1999-05-14
;  NUMBER OF SEQ ID NOS:  [illegible]
;  SOFTWARE:  PERL Program
; SEQ ID NO 5519
;   LENGTH: 291
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. 6476212.700350285H1
;   NAME/KEY: unsure
;   LOCATION: 6, 40, 54, 65, 67, 200, 221, 237, 238, 261
;   OTHER INFORMATION: a, t, c, g, or other
US-09-313-294A-5519

Alignment Scores:
Pred. No.:                2.95e-34        Length:          291
Score:                    383.50          Matches:         71
Percent Similarity:       82.66%          Conservative:    10
Best Local Similarity:    72.46%          Mismatches:      15
Query Match:              7.32%           Indels:          2
DB:                       4               Gaps:            1

US-09-980-043-2 (1-978) x US-09-313-294A-5519 (1-291)

Qy        687 ArgGlyTrpArgSerIleTyrCysMetProValArgProAlaPheLysGlySerAlaPro 706
              |||  |||||||||| |||:|||:|  |  |  || ||||||||  ||||||
Db          3 CGANGCTGGCGATCAATCTACTGCATGCCTAAGNGACNAGCTTTCAAGGGANCTGCTCCT 62

Qy        707 -IleAsnLeuSerAspArgLeuHisGlnValLeuArgTrpAlaLeuGlySerValGluIl 726
               | |||: |||||||||:::|||||:|||| ||||:|||||||,|:::|||||
Db         63 ATNCNACCTTTCGGATCGTTTGAATCAAGTGCTTCGGTGGGCTCTTGGTTCCATTGAAAT 122

Qy        726 ePhePheSerArgHisCysProLeuTrpTyrGlyPheGlyGlyGlyArgLeuLysTrpLe 746
              |   ||||||||||  ||||||:::|||||  |:::   ||||:|||| |||||:::||
Db        123 TCTTTTCAGCAGGCATTGTCCCATATGGTATCGTTAT---GGAGGTCGGCTTAAATTCCT 179

Qy        746 uGlnArgLeuAlaTyrIleAsnThrIleValTyrProPheThrSerLeuProLeuIleAl 766
              |:::|::  |||||:||  |||  :::|  |||       |||:::   |||:::
Db        180 GGAGAGATTTGCTTATATCANCAGAACAATTTATTCCACTCANATCAATCCNGCTCCTCNT 239

Qy        766 aTyrCysThrIleProAlaValCysLeuLeuThrGlyLysPheIleIlePro 783
              ||||||   ::: |||  |||:|||||: ||||: ||||| ||||| |||
Db        240 GTACTGCATATTGCAAGCAGTNTGTCTTCTTCACTGGGAAGTTCATCATCCCA 291

RESULT 14
US-09-313-294A-6485
; Sequence 6485, Application: US/09313294A
; Patent No. 6476212
; GENERAL INFORMATION:
;  APPLICANT:  Lalgudi, Raghunath V.
;  APPLICANT:  Ito, Laura Y.
;  APPLICANT:  Sherman, Bradley K.
;  TITLE OF INVENTION:  POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
;  FILE REFERENCE:  PL-0017 US
;  CURRENT APPLICATION NUMBER: US/09/313,294A
;  CURRENT FILING DATE:  1999-05-14
```

; NUMBER OF SEQ ID NOS: 7600
; SOFTWARE: PERL Program
; SEQ ID NO 6485
; LENGTH: 281
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. 6476212 7003518821H1
US-09-313-294A-6485

Alignment Scores:
Pred. No.: 1.51e-33 Length: 281
Score: 377.00 Matches: 66
Percent Similarity: 84.95% Conservative: 13
Best Local Similarity: 70.97% Mismatches: 14
Query Match: 7.19% Indels: 0
DB: 4 Gaps: 0

US-09-980-043-2 (1-978) x US-09-313-294A-6485 (1-281)

Qy 752 IleAsnThrIleValTyrProPheThrSerLeuProLeuIleAlaTyrCysThrIlePro 771
Db 2 ATCAACACCACCATCTACCCGTTCACGTCCCTCCCGCTCCTCATTTACTGTATCCTGCCT 61

Qy 772 AlaValCysLeuLeuThrGlyLysPheIleIleProThrLeuSerAsnLeuAlaSerMet 791
Db 62 GCCATCTGCCTGCTCACGGGGAAGTTCATCATCCCAGAGATCAGCAACTTCGCTAGTATC 121

Qy 792 LeuPheLeuGlyLeuPheIleSerIleIleValThrAlaValLeuGluLeuArgTrpSer 811
Db 122 TGGTTCATCTCTCTCTTCATCTCGATCTTCGCCACGGGTATCCTGGAGATGAGGTGGAGC 181

Qy 812 GlyValSerIleGluAspLeuTrpArgAsnGluGlnPheTrpValIleGlyGlyValSer 831
Db 182 GGCGTGGGCATCGACGAGTGGTGGAGGAACGAGCAGTTCTGGGTCATCGGAGGCATCTCC 241

Qy 832 AlaHisLeuPheAlaValPheGlnGlyPheLeuLysMet 844
Db 242 GCCCACCTCTTCGCCGTCTTCCAGGGCCTCCTCAAGGTG 280

RESULT 15
US-09-313-294A-4753
; Sequence 4753, Application US/09313294A
; Patent No. 6476212
; GENERAL INFORMATION:
; APPLICANT: Lalgudi, Raghunath V
; APPLICANT: Ito, Laura Y.
; APPLICANT: Sherman, Bradley K
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
; FILE REFERENCE: PL-0017 US
; CURRENT APPLICATION NUMBER: US/09/313,294A
; CURRENT FILING DATE: 1999-05-14
; NUMBER OF SEQ ID NOS: 7600
; SOFTWARE: PERL Program
; SEQ ID NO 4753
; LENGTH: 300
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. 6476212 7003489828H1
; NAME/KEY: unsure
; LOCATION: 64, 189
; OTHER INFORMATION: a, t, c, g, or other
US-09-313-294A-4753

Alignment Scores:
Pred. No.: 5.75e-32 Length: 300
Score: 363.50 Matches: 77
Percent Similarity: 84.85% Conservative: 11
Best Local Similarity: 71.72% Mismatches: 14
Query Match: 6.94% Indels: 2
DB: 4 Gaps: 1

US-09-980-043-2 (1-978) x US-09-313-294A-4753 (1-300)

Qy 785 LeuSerAsnLeuAlaSerMetLeuPheLeuGlyLeuPheIleSerIleIleValThrAla 804
Db 5 ATCAGCAACTTCGCCAGCATCTGGTTCATCTCCCTCTTCATCTCGATCTTCGCCACGGGN 64

Qy 805 ValLeuGluLeuArgTrpSerGlyValSerIleGluAspLeuTrpArgAsnGluGlnPhe 824
Db 65 ATCCTGGAGATGAGGTGGAGCGGGGTGGGCATCGACGAGTGGTGGAGGAACGAGCAGTTC 124

Qy 825 TrpValIleGlyGlyValSerAlaHisLeuPheAlaValPheGlnGlyPheLeuLysMet 844
Db 125 TGGGTGATCGGGGGCATCTCCGCGCACCTCTTCGCCGTGTTCCAGGGCCTGCTCAAGGTG 184

Qy 845 LeuAlaGlyIleAspThrAsnPheThrValThrAlaLysAlaAla---GluAspAlaGlu 863
Db 185 CTGGNCGGCATCGACACCAACTTCACCGTCACCTCCAAGGCCTCGGACGAGGACGGCGAC 244

Qy 864 PheGlyGluLeuTyrMetValLysTrpThrThrLeuLeuIleProProThrThrLeu 882
Db 245 TTCGCGGA-CTGTACATGTTCAAGTGGACGACGCTCCTTATCCCGCCCACCACCATC 300

Search completed: October 31, 2003, 17:45:30
Job time : 243.118 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 31, 2003, 03:25:50 ; Search time 729.785 Seconds
(without alignments)
7667.830 Million cell updates/sec

Title: US-09-980-043-2
Perfect score: 5240
Sequence: 1 MMESGAPICHTCGEQVGHDA.....PPVNKVICTLVAETCISIDC 978

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
Q:/cgn2_1/USPTO_spool/US09980043/runat_29102003_163852_21851/app_query.fasta_1.1358
-DB=N_Geneseq_19Jun03 -QFMT=fastap_SUFFIX=rng -MINMATCH=0.1 -LOOPCL=0
LOOPEXT=0 UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 TRANS=human40.cdi
LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09980043_@CGN_1_1_458_@runat_29102003_163852_21851 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqn-eml/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-eml/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-eml/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-eml/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-eml/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-eml/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-eml/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-eml/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-eml/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-eml/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-eml/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-eml/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-eml/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-eml/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-eml/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-eml/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-eml/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-eml/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-eml/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-eml/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-eml/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-eml/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-eml/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-eml/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseqn-eml/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	5240	100.0	3232	22	AAC65448	Populus tremuloide
2	4548	86.8	3207	19	AAV08372	Cellulose synthase
3	4540	86.6	3328	19	AAV34432	Cotton cellulose s
4	4368	83.4	3747	21	AAA67145	Eucalyptus grandis
5	3492.5	66.7	3311	19	AAV08373	Cellulose synthase
6	3474	66.3	3851	21	AAA67114	Pinus radiata cell
7	3366.5	64.2	3799	24	AAS16458	Corn cDNA encoding
8	3365.5	64.2	3746	21	AAZ99512	DNA encoding a mai
9	3365.5	64.2	3746	21	AAZ99527	DNA encoding a mai
10	3365.5	64.2	3773	21	AAZ99494	DNA encoding a mai
11	3363	64.2	3776	21	AAZ58263	Corn cellulose syn
12	3362	64.2	3517	21	AAZ58268	Soybean cellulose
13	3356.5	64.1	3704	21	AAZ99533	DNA encoding a mai
14	3353	64.0	3568	21	AAZ99491	DNA encoding a mai
15	3350	63.9	3198	21	AAC49550	Arabidopsis thalia
16	3350	63.9	3603	19	AAV06565	Arabidopsis cellul
17	3345.5	63.8	3936	21	AAZ58266	Corn cellulose syn
18	3345.5	63.8	3969	21	AAZ99503	DNA encoding a mai
19	3345.5	63.8	3969	21	AAZ99518	DNA encoding a mai
20	3345	63.8	3786	21	AAZ58265	Corn cellulose syn
21	3345	63.8	3813	21	AAZ99509	DNA encoding a mai
22	3345	63.8	3813	21	AAZ99524	DNA encoding a mai
23	3340	63.7	3673	19	AAV06568	Arabidopsis cellul
24	3335.5	63.7	3753	21	AAZ99515	DNA encoding a mai
25	3335.5	63.7	3753	21	AAZ99530	DNA encoding a mai
26	3335.5	63.7	3780	21	AAZ99497	DNA encoding a mai
27	3332	63.6	3725	21	AAZ99500	DNA encoding a mai
28	3332	63.6	3725	21	AAZ99506	DNA encoding a mai
29	3332	63.6	3725	21	AAZ99521	DNA encoding a mai
30	3329	63.5	3614	19	AAV06567	Arabidopsis cellul
31	3274	62.5	3255	24	ABZ12754	Arabidopsis thalia
32	3242	61.9	3828	19	AAV06566	Arabidopsis cellul
33	3240.5	61.8	3255	24	ABZ14725	Arabidopsis thalia
34	3213.5	61.3	3444	22	AAC83798	Arabidopsis thalia
35	3156	60.2	2830	24	AAS16455	Corn cDNA encoding
36	3070	58.6	2890	21	AAZ58270	Soybean cellulose
37	3042	58.1	7234	22	AAC83247	Cellulose synthase
38	2925.5	55.8	8411	19	AAV06563	Arabidopsis cellul
39	2734	52.2	2306	19	AAV34433	Cotton cellulose s
40	2435.5	46.5	2033	19	AAV08381	Cellulose synthase
41	2311	44.1	5009	19	AAV06564	Arabidopsis cellul
42	2286.5	43.6	2055	21	AAA67111	Eucalyptus grandis
43	2284.5	43.6	2248	19	AAV06562	Arabidopsis cellul
44	2260	43.1	2125	21	AAZ58269	Soybean cellulose
45	2180.5	41.6	1407	21	AAA67086	Eucalyptus grandis

ALIGNMENTS

RESULT 1
AAC65448
ID AAC65448 standard; cDNA; 3232 BP.
XX
AC AAC65448;
XX
DT 07-MAR-2001 (first entry)
XX
DE Populus tremuloides CelA cDNA.
XX
KW Aspen; CelA; cellulose synthase; arabidopsis; wood;
KW mechanical stress response element; MSRE; cellulose biosynthesis; ss.
XX
OS Populus tremuloides.
XX
PN WO200071670-A2.
XX

30-NOV-2000.

18-MAY-2000; 2000WO-US13637.

21-MAY-1999; 99US-0135280.

(UNMT) UNIV MICHIGAN TECHNOLOGICAL.

Chiang VL, Wu L, Chandrashekhar JP, Carraway DT;

WPI; 2001-032025/04.
P-PSDB; AAB37893.

Novel polynucleotide encoding cellulose synthase useful for enhancing cellulose biosynthesis, modifying lignin biosynthesis, and for accelerating growth of plants and for upregulating cellulose synthase levels -

Claim 1; Fig 1; 65pp; English.

The present sequence is given in a specification relating to polynucleotides encoding cellulose synthase enzymes from Populus tremuloides and Arabidopsis. The cellulose synthases are designated PtCelA and AraxCelA respectively. The polynucleotides may be introduced into plants where they are useful for altering the growth of a plant. Vectors comprising a cellulose synthase promoter, or a fragment of the promoter containing one or more mechanical stress response elements (MSRE), operably linked to a cellulose synthase gene are useful for causing stress-induced gene expression in a plant cell. A cDNA encoding a protein that binds to a positive MSRE may be delivered into the plant to increase expression of cellulose synthase. A cDNA in an antisense orientation may be delivered to reduce expression of cellulose synthase. The polynucleotides enable the development of transgenic tree varieties having increased cellulose content, decreased lignin content and therefore improved wood fibre characteristics. The cellulose synthase promoters are useful for synthesising greater quantities of high-crystalline cellulose in plant and preferably in trees, which permits an increase in proportion of cellulose in transgenic plants, greater strength of juvenile wood and fiber and acceleration of overall growth rate.

Sequence 3232 BP; 895 A; 668 C; 772 G; 897 T; 0 other;

lignment Scores:

red. No.:	0	Length:	3232
core:	5240.00	Matches:	978
ercent Similarity:	100.00%	Conservative:	0
est Local Similarity:	100.00%	Mismatches:	0
uery Match:	100.00%	Indels:	0
B:	22	Gaps:	0

S-09-980-043-2 (1-978) x AAC65448 (1-3232)

```
Qy          1 MetMetGluSerGlyAlaProIleCysHisThrCysGlyGluGlnValGlyHisAspAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         69 ATGATGGAATCTGGGGCTCCTATATGCCATACCTGTGGTGAACAGGTGGGGCATGATGCA 128

Qy         21 AsnGlyGluLeuPheValAlaCysHisGluCysSerTyrProMetCysLysSerCysPhe 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        129 AATGGGGAGCTATTTGTGGCTTGCCATGAGTGTAGCTATCCCATGTGCAAGTCTTGTTTC 188

Qy         41 GluPheGluIleAsnGluGlyArgLysValCysLeuArgCysGlySerProTyrAspGlu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        189 GAGTTTGAAATCAATGAGGGCCGGAAAGTTTGCTTGCGGTGTGGCTCGCCATATGATGAG 248

Qy         61 AsnLeuLeuAspAspValGluLysLysGlySerGlyAsnGlnSerThrMetAlaSerHis 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        249 AACTTGCTGGATGATGTAGAAAAGAAGGGGTCTGGCAATCAATCCACAATGGCATCTCAC 308

Qy         81 LeuAsnAspSerGlnAspValGlyIleHisAlaArgHisIleSerSerValSerThrVal 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        309 CTCAACGATTCTCAGGATGTCGGAATCCATGCTAGACATATCAGTAGTGTGTCCACTGTG 368

Qy        101 AspSerGluMetAs AspGluTyrGlyAsnProIleTrpLysAsnArgValLysSerCys 120
              |||||||||||||| ||| |||| |||||||||||||||||||||||||||||||||||||
Db        369 GATAGTGAAATGAA GATGAATATGGGAATCCAATTTGGAAGAATCGGGTGAAGAGCTGT 428

Qy        121 LysAspLysGluAs LysLysLysLysArgSerProLysAlaGluThrGluProAlaGln 140
              |||||||||||||| |||||||| ||| |||||||||||||||||||||||||||||||||||
Db        429 AAGGATAAAGAGAA AAGAAGAAAAAGAGAAGTCCTAAGGCTGAAACTGAACCAGCTCAA 488

Qy        141 ValProThrGluGl GlnMetGluGluLysProSerAlaGluAlaSerGluProLeuSer 160
              |||||||||||||| |||||||| |||||||||||||||||||||||||||||||||||||||
Db        489 GTTCCTACAGAACA AAGATGGAAGAGAAACCGTCTGCAGAGGCTTCGGAGCCGCTTTCA 548

Qy        161 IleValTyrProIl ProArgAsnLysLeuThrProTyrArgAlaValIleIleMetArg 180
              |||||||||||||| |||||||| |||||||||||||||||||||||||||||||||||||||
Db        549 ATTGTTTATCCAAT CCAGGAACAAGCTCACACCATACAGAGCAGTGATCATTATGCGA 608

Qy        181 LeuValIleLeuT ValLeuHisGlyIleArgIleThrAsnProValAspSerAlaPhe 200
              ||| ||| ||| || ||| |||||| ||| |||||||||||||||||||||||||||||||||
Db        609 CTGGTCATTCTGT TCTTTCATTATAGAATAAGAAATCCTGTCGATAGTGCCTTT 668

Qy        201 GlyLeuTrpLeuT SerValIleCysGluIleTrpPheAlaPheSerTrpValLeuAsp 220
              |||| |||| ||| |||||||| ||| ||||||||||||||||||||||||||||||||||||
Db        669 GGCCTGTGGCTTAC TCTGTCATATGTGAGATCTGGTTTGCATTTTCTTGGGTGTTGGAT 728

Qy        221 GlnPheProLysTr AsnProValAsnArgGluThrTyrIleGluArgLeuSerAlaArg 240
              |||||||||||||| |||||||| |||||||||||||||||||||||||||||||||||||||
Db        729 CAGTTCCCCAAGTG AATCCTGTCAATAGAGAAACGTATATCGAAAGGCTGTCGGCAACG 788

Qy        241 TyrGluArgGluGl GluProSerGlnLeuAlaGlyValAspPhePheValSerThrVal 260
              |||||||||||||| |||||||| |||||||||||||||||||||||||||||||||||||||
Db        789 TATGAAAGAGAGGG GAGCCTTCTCAGCTTGCTGGTGTGGATTTTTTCGTGAGTACTGTT 848

Qy        261 AspProLeuLysGl ProProLeuIleThrAlaAsnThrValLeuSerIleLeuAlaVal 280
              |||||||||||||| |||||||| |||||||||||||||||||||||||||||||||||||||
Db        849 GATCCGCTGAAGGAACCGCCATTGATCACTGCAAATACAGTCCTTTCCATCCTTGCTGTG 908

Qy        281 AspTyrProValAs LysValSerCysTyrValSerAspAspGlyAlaAlaMetLeuSer 300
              |||||||| |||| |||| || |||| |||| ||||||||||||||||||||||||||||||||
Db        909 GACTATCCCGTCGA AAAGTCTCCTGCTACGTGTCTGATGATGGTGCAGCTATGCTTTCA 968

Qy        301 PheGluSerLeuVa GluThrAlaGluPheAlaArgLysTrpValProPheCysLysLys 320
              |||||||||||||| ||| |||| |||||||||||||||||||||||||||||||||||||||
Db        969 TTTGAATCTCTTGTGAAAACAGCTGAGTTTGCAAGGAAGTGGGTTCCGTTCTGCAAAAAA 1028

Qy        321 PheSerIleGluPr ArgAlaProGluPheTyrPheSerGlnLysIleAspTyrLeuLys 340
              |||||||||||||| |||||||| |||||||||||||||||||||||||||||||||||||||
Db       1029 TTCTCAATTGAACCAAGAGCACCGGAGTTTTACTTCTCACAGAAAATTGATTACTTGAAA 1088

Qy        341 AspLysValGlnPr SerPheValLysGluArgArgAlaMetLysArgAspTyrGluGlu 360
              |||||||||||||| |||||||| |||||||||||||||||||||||||||||||||||||||
Db       1089 GACAAGGTTCAACC TCTTTCGTGAAAGAACGTAGAGCAATGAAAAGGGATTATGAAGAG 1148

Qy        361 TyrLysValArgVa AsnAlaLeuValAlaLysAlaGlnLysThrProGluGluGlyTrp 380
              |||||||||||||| |||||||| |||||||||||||||||||||||||||||||||||||||
Db       1149 TACAAAGTCCGAGT AATGCCCTGGTAGCAAAGGCTCAGAAAACACCTGAAGAAGGATGG 1208

Qy        381 ThrMetGlnAspGl ThrProTrpProGlyAsnAsnThrArgAspHisProGlyHisAsp 400
              |||||||||||||| |||||||| |||||||||||||||||||||||||||||||||||||||
Db       1209 ACTATGCAAGATGG AACACCTTGGCCTGGGAATAACACACGTGATCACCCTGGGCATGAT 1268

Qy        401 SerGlyLeuProTr GluIleLeuGlyAlaArgAspIleGluGlyAsnGluLeuProArg 420
              |||||||||||||| |||||||| |||||||||||||||||||||||||||||||||||||||
Db       1269 TCAGGTCTTCCTTG GAAATACTGGGAGCTCGTGACATTGAAGGAAATGAACTACCTCGT 1328

Qy        421 LeuValTyrValSe ArgGluLysArgProGlyTyrGlnHisHisLysLysAlaGlyAla 440
              |||||||||||||| |||||||| |||||||||||||||||||||||||||||||||||||||
Db       1329 CTAGTATATGTCTC AGGGAGAAGAGACCTGGCTACCAGCACCACAAAAAGGCTGGTGCA 1388

Qy        441 GluAsnAlaLeuValArgValSerAlaValLeuThrAsnAlaProTyrIleLeuAsnVal 460
              |||||||||||||| |||||||| |||||||||||||||||||||||||||||||||||||||
Db       1389 GAAAATGCTCTGGTGAGAGTGTCTGCAGTACTCACAAATGCTCCCTACATCCTCAATGTT 1448

Qy        461 AspCysAspHisTyrValAsnAsnSerLysAlaValArgGluAlaMetCysIleLeuMet 480
```

```
Db       1449 GATTGTGATCACTATGTAAACAATAGCAAGGCTGTTCGAGAGGCAATGTGCATCCTGATG 1508

Qy        481 AspProGlnValGlyArgAspValCysTyrValGlnPheProGlnArgPheAspGlyIle 500

Db       1509 GACCCACAAGTAGGTCGAGATGTATGCTATGTGCAGTTCCCTCAGAGGTTTGATGGCATA 1568

Qy        501 AspLysSerAspArgTyrAlaAsnArgAsnValValPheProAspValAsnMetLysGly 520

Db       1569 GATAAGAGTGATCGCTACGCCAATCGTAACGTAGTTTTCTTTGATGTTAACATGAAAGGG 1628

Qy        521 LeuAspGlyIleGlnGlyProValTyrValGlyThrGlyCysValPheAsnArgGlnAla 540

Db       1629 TTGGATGGCATTCAAGGACCAGTATACGTAGGAACTGGTTCTGTTTTCAACAGGCAAGCA 1688

Qy        541 LeuTyrGlyTyrGlyProProSerMetProSerLeuArgLysArgLysAspSerSerSer 560

Db       1689 CTTTACGGCTACGGGCCTCCTTCTATGCCCAGCTTACGCAAAGAAAGGATTCTTCATCC 1748

Qy        561 CysPheSerCysCysCysProSerLysLysLysProAlaGlnAspProAlaGluValTyr 580

Db       1749 TGCTTCTCATGTTGCTGCCCCTCAAAGAAGAAGCCTGCTCAAGATCCAGCTGAGGTATAC 1808

Qy        581 ArgAspAlaLysArgGluAspLeuAsnAlaAlaIlePheAsnLeuThrGluIleAspAsn 600

Db       1809 AGAGATGCAAAAAGAGAGGATTTCAATGCTGCCATATTTAATCTTACAGAGATTGATAAT 1868

Qy        601 TyrAspGluHisGluArgSerMetLeuIleSerGlnLeuSerPheGluLysThrPheGly 620

Db       1869 TATGACGAGCATGAAAGGTCAATGCTGATCTCCCAGTTGAGCTTTGAGAAAACTTTTGGC 1928

Qy        621 LeuSerSerValPheIleGluSerThrLeuMetGluAsnGlyGlyValProGluSerAla 640

Db       1929 TTATCTTCTGTCTTCATTGAGTCTACACTAATGGAGAATGGAGGAGTACCCGAGTCTGCC 1988

Qy        641 AsnSerProProPheIleLysGluAlaIleGlnValIleGlyCysGlyTyrGluGluLys 660

Db       1989 AACTCACCACCATTCATCAAGGAAGCGATTCAAGTCATCGGCTGTGGCTATGAAGAGAAG 2048

Qy        661 ThrGluTrpGlyLysGlnIleGlyTrpIleTyrGlySerValThrGluAspIleLeuSer 680

Db       2049 ACTGAATGGGGAAAACAGATTGGTTGGATATATGGGTCAGTCACTGAGGATATCTTAAGT 2108

Qy        681 GlyPheLysMetHisCysArgGlyTrpArgSerIleTyrCysMetProValArgProAla 700

Db       2109 GGCTTCAAGATGCACTGCCGAGGATGGAGATCAATTTACTGTATGCCCGTAAGGCCTGCA 2168

Qy        701 PheLysGlySerAlaProIleAsnLeuSerAspArgLeuHisGlnValLeuArgTrpAla 720

Db       2169 TTCAAAGGATCTGCACCCATCAACCTGTCTGATAGACTGCACCAGGTCCTCCGATGGGCT 2228

Qy        721 LeuGlySerValGluIlePheLeuSerArgHisCysProLeuTrpTyrGlyPheGlyGly 740

Db       2229 CTTGGTTCTGTGGAAATTTTCTTTAGCAGACACTGTCCCCTTGGTACGGGTTTGGAGGA 2288

Qy        741 GlyArgLeuLysTrpLeuGlnArgLeuAlaTyrIleAsnThrIleValTyrProPheThr 760

Db       2289 GGCCGTCTTAAATGGCTCCAAAGGCTTGCGTATATAAACACTATTGTGTACCCATTTACA 2348

Qy        761 SerLeuProLeuIleAlaTyrCysThrIleProAlaValCysLeuLeuThrGlyLysPhe 780

Db       2349 TCCCTCCCTCTCATTGCCTATTGCACAATTCCTGCAGTTTGTCTGCTCACCGGAAAATTC 2408

Qy        781 IleIleProThrLeuSerAsnLeuAlaSerMetLeuPheLeuGlyLeuPheIleSerIle 800

Db       2409 ATCATACCAACGCTCTCAAACCTGGCAAGCATGCTGTTTCTTGGCCTCTTTATCTCCATC 2468

Qy        801 IleValThrAlaValLeuGluLeuArgTrpSerGlyValSerIleGluAspLeuTrpArg 820

Db       2469 ATTGTAACTGCGGTGCTTGAGCTAAGATGGAGCGGTGTCAGCATTGAAGATTTATGGCGT 2528

Qy        821 AsnGluGlnPheTrpValIleGlyGlyValSerAlaHisLeuPheAlaValPheGlnGly 840

Db       2529 AATGAACAATTCTGGGTGATCGGAGGTGTTTCAGCCCATCTCTTTGCGGTCTTCCAGGGA 2588

Qy        841 PheLeuLysMetLeuAlaGlyIleAspThrAsnPheThrValThrAlaLysAlaAlaGlu 860

Db       2589 TTCTTAAAAATGTTGGCTGGCATCGATACGAACTTCACTGTCACAGCAAAAGCAGCCGAA 2648

Qy        861 AspAlaGluPheGlyGluLeuTyrMetValLysTrpThrThrLeuLeuIleProProThr 880

Db       2649 GATGCAGAATTTGGGGAGCTATATATGGTCAAGTGGACAACACTTTTGATTCCTCCAACC 2708

Qy        881 ThrLeuLeuIleIleAsnMetSerGlyCysAlaGlyPheSerAspAlaLeuAsnLysGly 900

Db       2709 ACACTTCTCATTATCAATATGTCGGGTTGTGCTGGATTCTCTGATGCACTCAACAAAGGA 2768

Qy        901 TyrGluAlaTrpGlyProLeuPheGlyLysValPhePheAlaPheTrpValIleLeuHis 920

Db       2769 TATGAAGCATGGGGGCCTCTCTTTGGCAAGGTGTTCTTTGCTTTCTGGGTGATTCTTCAT 2828

Qy        921 LeuTyrProPheLeuLysGlyLeuMetGlyArgGlnAsnLeuThrProThrIleValVal 940

Db       2829 CTCTATCCATTCCTTAAAGGTCTAATGGGTCGCCAAAACCTAACACCAACCATTGTTGTT 2888

Qy        941 LeuTrpSerValLeuLeuAlaSerValPheSerLeuValTrpValLysIleAsnProPhe 960

Db       2889 CTCTGGTCAGTGCTGTTGGCCTCTGTCTTCTCTCTCGTTTGGGTCAAGATCAATCCATTC 2948

Qy        961 ValAsnLysValAspAsnThrLeuValAlaGluThrCysIleSerIleAspCys 978

Db       2949 GTTAACAAAGTTGATAACACCTTGGTTGCGGAGACCTGCATTTCCATTGATTGC 3002
```

RESULT 2
AAV08372
ID AAV08372 standard; cDNA to mRNA; 3207 BP.
XX
AC AAV08372;
XX
DT 08-FEB-1999 (first entry)
XX
DE Cellulose synthase PcsA1 coding sequence.
XX
KW Cellulose synthase; cotton; ds.
XX
OS Gossypium hirsutum.
XX
FH Key Location/Qualifiers
FT CDS 77..3001
FT /*tag= a
XX
PN EP875575-A2.
XX
PD 04-NOV-1998.
XX
PF 31-MAR-1998; 98EP-0302489.
XX
PR 01-APR-1997; 97JP-0083133.
XX
PA (NISN) NISSHINBO IND INC.
XX
PI Aotsuka S, Hasegawa O, Hayashi T, Ihara Y;
XX
DR WPI; 1998-559440/48.
DR P-PSDB; AAW73308.
XX
PT New DNA encoding cotton cellulose synthase proteins - useful for
PT controlling cellulose synthesis in prokaryotic or eukaryotic cells
XX
PS Example 5; Page 14-18; 51pp; English.
XX
CC This sequence encodes a cellulose synthase of the invention. The DNA can
CC be used for controlling cellulose synthesis in a cell by introducing the
CC DNA into the cell, expressing RNA with a nucleotide sequence homologous
CC to the DNA or a nucleotide sequence complementary to the DNA. The cells
CC are prokaryotic cells i.e. acetobacterium and/or eukaryotic cells such as

yeasts i.e. Saccharomyces spp., plant cells (cotton plants) and cultured cells of mammals etc.

Sequence 3207 BP; 883 A; 662 C; 723 G; 939 T; 0 other;

Alignment Scores:

Seq. No.:	0	Length:	3207
Score:	4548.00	Matches:	845
Percent Similarity:	93.47%	Conservative:	71
Best Local Similarity:	86.22%	Mismatches:	57
Query Match:	86.79%	Indels:	8
DB:	19	Gaps:	6

US-09-980-043-2 (1-978) x AAV08372 (1-3207)

```
Qy          1 MetMetGluSerGlyAlaProIleCysHisThrCysGlyGluGlnValGlyHisAspAla 20
              ||||||||||||||||   |||:::|||||||||||||||||||   ||||||   :::
Db         77 ATGATGGAATCTGGGGTTCCTGTTTGCCACACTTGTGGTGAACATGTTGGGTTGAATGTT 136

Qy         21 AsnGlyGluLeuPheValAlaCysHisGluCysSerTyrProMetCysLysSerCysPhe 40
              |||||||||   |||||||||||||||||||||||:::::|||:::||||||||||||||
Db        137 AATGGTGAACCCTTTGTGGCTTGCCATGAATGTAATTTCCCTATTTGTAAGAGTTGTTTT 196

Qy         41 GluPheGluIleAsnGluGlyArgLysValCysLeuArgCysGlySerProTyrAspGlu 60
              |||:::::::::   |||||||:::|||   |||||||||||||||   ||||||||||||
Db        197 GAGTATGATCTTAAGGAAGGACAAAAAGCTTGCTTGCGTTGTGGTATTCCGTATGATGAA 256

Qy         61 AsnLeuLeuAspAspValGluLysLysGlySerGlyAsnGlnSerThrMetAlaSerHis 80
              ||||||||||||||||||||   |||   :::|||:::|||||||||||||||||:::|||
Db        257 AACCTGTTGGACGATGTCGAG---AAGGCCACCGGCGATCAATCGACAATGGCTGCACAT 313

Qy         81 LeuAsnAspSerGlnAspValGlyIleHisAlaArgHisIleSerSerValSerThrVal 100
              |||:::   ||||||||||||||||||||||||||||||||||||||||||||||:::
Db        314 TTGAGCAAGTCTCAGGATGTTGGAATTCATGCAAGACATATCAGCAGTGTGTCTACATTG 373

Qy        101 AspSerGluMetAsnAspGluTyrGlyAsnProIleTrpLysAsnArgValLysSerCys 120
              ||||||||||||   ::::::   ||||||||||||||||||||||||:::|||
Db        374 GATAGTGAAATGACTGAAGACAATGGGAATCCGATTTGGAAGAACAGGGTGGAAAGTTGG 433

Qy        121 LysAspLysGluAsnLysLysLysLysArgSerProLysAlaGluThrGluProAlaGln 140
              |||:::|||:::|||||||||||||||   :::                   ||||:::
Db        434 AAAGAAAAGAAGAACAAGAAGAAGAAGCCTGCAACAACTAAGGTTGAAAGAGAGGCTGAA 493

Qy        141 ValProThrGluGlnGlnMetGluGluLysProSerAlaGluAlaSerGluProLeuSer 160
              :::|||   ||||||||||||||||:::|||||||:::   :::|||||||:::||||||||
Db        494 ATCCCACCTGAGCAACAAATGGAAGATAAACCGGCACCGGATGCTTCCCAGCCCCTCTCG 553

Qy        161 IleValTyrProIleProArgAsnLysLeuThrProTyrArgAlaValIleIleMetArg 180
              :::   |||||||||:::::::|||   |||||||||   ||||||||||||||||||||
Db        554 ACTATAATTCCAATCCCGAAAAGCAGACTTGCACCATACCGAACCGTGATCATTATGCGA 613

Qy        181 LeuValIleLeuGlyLeuPhePheHisPheArgIleThrAsnProValAspSerAlaPhe 200
              |||:::|||||||||||||||||||:::||:::|||||||||||||||||||||||||||
Db        614 TTGATCATTCTCGGTCTTTTCTTCCATTATCGAGTAACAAACCCCGTTGACAGTGCTTTT 673

Qy        201 GlyLeuTrpLeuThrSerValIleCysGluIleTrpPheAlaPheSerTrpValLeuAsp 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        674 GGACTGTGGCTCACTTCAGTCATATGTGAAATCTGGTTTGCTTTTTCCTGGGTGTTGGAT 733

Qy        221 GlnPheProLysTrpAsnProValAsnArgGluThrTyrIleGluArgLeuSerAlaArg 240
              ||||||||||||||   ||||||||||||||||||:::|||||||||||||||||||||
Db        734 CAGTTCCCTAAGTGGTATCCTGTTAACAGGGAAACATACATTGACAGACTGTCTGCAAGA 793

Qy        241 TyrGluArgGluGlyGluProSerGlnLeuAlaGlyValAspPhePheValSerThrVal 260
              ||||||||||||||||||||||:::::|||||   ||||||||||||||||||||||||
Db        794 TATGAAAGAGAAGGTGAACCTAATGAACTTGCTGCAGTTGACTTCTTTGTGAGTACAGTG 853

Qy        261 AspProLeuLysGluProProLeuIleThrAlaAsnThrValLeuSerIleLeuAlaVal 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db        854 GATCCATTGAAAGAGCCTCCATTGATTACTGCCAATACTGTGCTTTCCATCCTTGCCTTG 913

Qy        281 AspTyrProValAspLysValSerCysTyrValSerAspAspGlyAlaAlaMetLeuSer 300
              ||||||||||||||||||||||||||||||:::|||||||||||||||||||||||:::
```

```
Db        914 GACTACCCGGTAGA?AAGGTCTCTTGTTATATATCTGATGATGGTGCGGCCATGCTGACA 973

Qy        301 PheGluSerLeuVa?GluThrAlaGluPheAlaArgLysTrpValProPheCysLysLys 320
              |||||||||||||:|||.|||:::::| |||||:|||||||||||||||||||||||||
Db        974 TTTGAATCTCTAGT?GAAACAGCCGACTTTGCAAGAAAGTGGGTTCCATTCTGCAAAAAA 1033

Qy        321 PheSerIleGluPr ArgAlaProGluPheTyrPheSerGlnLysIleAspTyrLeuLys 340
              ||||||||||:|||. |||:|||| |||:|||||||||||||||||||||||||||||||
Db       1034 TTTTCCATTGAAC?A?GGCACCTGATGTTTACTTCTCACAGAAGATTGATTACTTGAAA 1093

Qy        341 AspLysValGlnPr SerPheValLysGluArgArgAlaMetLysArgAspTyrGluGlu 360
              ||||'||||:|||  |||  |||  ||||:|||||||||||||||||||||||||||||||
Db       1094 GATAAAGTGCAGCC ?CTTTGTAAAAGAACGTAGAGCTATGAAAAGAGATTACGAAGAG 1153

Qy        361 TyrLysValArgVa AsnAlaGlnValAlaLysAlaGlnLysThrProGluGluGlyTrp 380
              |||||:::|:|:: :  |||  '||.  |||   ||||:|||||||||  ||||:
Db       1154 TACAAAATTCGAAT ?AA?CTTCA?TT? AAA ?? TCAGAAAACACCTGAAGAAGGATGG 1213

Qy        381 ThrMetGluAsp[ ..:|:::|:|:: AyA..ArgThrArgAspHisProGlyHisAsp 400
              |||'|||  |||:|  |||  :|:|  |||||||||||||||||||||||||  |||||
Db       1214 ACAATGCAAGATG?A?TCCTTGGCCA?AAA?AACCGCGTGATCACCCTGG-CATGAT 1272

Qy        401 SerGlyLeuProTr GluIle LeuGlyAlaArgAspIleGluGlyAsnGluLeuProAr 420
              ||||||  ||||:|  |||  ||.||  :|||||||||||||||:|||||:|
Db       1273 TCAGGTTTTCCTTG?---ATATAGCGGTGCTCATGACATCGAAGGAAATGAACTTCCCCG 1329

Qy        421 gLeuValTyrValSer ArgGluLysArgProGlyTyrGlnHisHisLysLysAlaGlyAl 440
              ||||||||||:|||  |||  |||||||||  ||||:|||||||||||||||||:||||
Db       1330 ACTGGTTTACGTCT?TAGAGAGAAGAGACCTGGCTACCAACACCACAAAAAGGCTGGTGC 1389

Qy        441 aGluAsnAlaLeuVa?ArgValSerAlaValLeuThrAsnAlaProTyrIleLeuAsnVa 460
              |||||||||||||:|||||||||||||:|||||:|||||||||||||||||||||||||
Db       1390 TGAAAATGCTTTGGT?AGGGTGTCTGCAGTTCTTACAAATGCTCCCTTCATCCTCAATCT 1449

Qy        461 lAspCysAspHisTyrValAsnAsnSerLysAlaValArgGluAlaMetCysIleLeuMe 480
              :||||||||||||| .|||:|||:|||| ||||||||||||||||||||||||||:|||.
Db       1450 TGATTGTGACCACTATGTTAACAATAGCAAGGCAGTTAGGGAGGCAATGTGCTTCTTGAT 1509

Qy        481 tAspProGlnValGlyArgAspValCysTyrVa?GlnPheProGlnArgPheAspGlyIl 500
              |||||||||||||||||||||||||||||||:|||||||||||||||||||||||||||
Db       1510 GGACCCACAAGTCG?TCGAGATCTCTGCTATGTGCAGTTTCCTCAAAGATTTGATGGCAT 1569

Qy        501 eAspLysSerAspArgTyrAlaAsnArgAsnValValPhePheAspValAsnMetLysGl 520
              |||:::|||||||||:|||||||||:|||  |||||||||||||||||||||||||||||
Db       1570 AGATAGGAGTGATC?ATATGCCAATCGGAACACAGTTTTCTTTGATGTTAACATGAAAGG 1629

Qy        521 yLeuAspGlyIleGlnGlyProValTyrValGlyThrGlyCysValPheAsnArgGlnAl 540
              ||||||||||||' |||'||||||||||||.|:|||||||||||||||||||||||||||
Db       1630 TCTTGATGGAATC?AAGGGCCTGTTTATGTGGGAACAGGTTGTGTTTTCAATAGGCAAGC 1689

Qy        541 aLeuTyrGlyTyrG.yProProSerMetProSerLeuArgLysArgLysAspSerSerSe 560
              ||||||||||||||.|||||||||||||||||||||||||||||||||||||||||||||
Db       1690 ACTTTATGGCTATG?TCCACCTTCAATG?CAA?TTTCCCAAG-- TCATCCTCCTCATC 1746

Qy        561 rCysPheSerCysCysCysProSerLysLysLysProAlaGlnAspProAlaGluValTy 580
              ||||  ||||||  ||||:|  ||||  :|||  :::||||:|:::|||||||||||||
Db       1747 TTGC---TCGTGTT?CTGCCCCGGCAAGAAGGAA?CT---AAAGATCCATCAGAGCTTTA 1803

Qy        581 rArgAspAlaLysArgGluAspLeuAsnAlaAlaIlePheAsnLeuThrGluIleAspAs 600
              |||||||||||||||:|||:|||||||||||||||||||||||||||||||||||||||
Db       1804 TAGGGATGCAAAAC?GGAAGAACTTGATGCTG?CATCTTTAACCTTAGGGAAATTGACAA 1863

Qy        601 nTyrAspGluHisGluArgSerMetLeuIleSerGlnLeuSerPheGluLysThrPheGl 620
              |||||||:|||:|||||||||||||||||||||||||||||||||||||||||||||||
Db       1864 TTATGATGAGTATGAAAGATCAATGTTGATCTCTCAAACAAGCTTTGAGAAAACTTTTGG 1923

Qy        621 yLeuSerSerValPheIleGluSerThrLeuMetGluAsnGlyGlyValProGluSerAl 640
              ||||||||||||||:|||||||||||||||||||||||||||||||||||||||||||||
Db       1924 CTTATCTTCAGTCTTCATTGAATCTACACTAATGGAGAATGGAGGAGTGGCTGAATCTGC 1983

Qy        641 aAsnSerProProPheIleLysGluAlaIleGlnValIleGlyCysGlyTyrGluGluLy 660
              ||||  ||||||||||||||||  |||||||||||||||||||||||||||||||||||
Db       1984 CAACCCTTCCACACTAATCAAGGAAGCAATTCATGTCATCGGCTGTGGCTATGAGGAGAA 2043
```

```
Qy        660 sThrGluTrpGlyLysGlnIleGlyTrpIleTyrGlySerValThrGluAspIleLeuSe 680
              ||||   |||||||||:::|.|||||||||||||||||||||||||||||||||||||||::
Db       2041 GACTGCATGGGGGAAAGAGATTGGATGGATATATGGTTCAGTCACTGAGGATATCTTAAC 2100

Qy        680 rGlyPheLysMetHisCysArgGlyTrpArgSerIleTyrCysMetProValArgProAl 700
              :|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 CGGCTTCAAAATGCACTGCCGAGGATGGAGATCGATTTACTGCATGCCCTTAAGGCCAGC 2160

Qy        700 aPheLysGlySerAlaProIleAsnLeuSerAspArgLeuHisGlnValLeuArgTrpAl 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 ATTCAAAGGATCTGCACCCATCAATCTGTCTGATCGGTTGCACCAGGTTCTTCGATGGGC 2220

Qy        720 aLeuGlySerValGluIlePhePheSerArgHisCysProLeuTrpTyrGlyPheGlyGl 740
              ||||||||||||||||||||  ||||||||||||||||||||||||||||||||||||||||
Db       2221 TCTTGGATCTGTTGAAATTTTCCTAAGCAGGCATTGCCCTCTATGGTATGGCTTTGGAGG 2280

Qy        740 yGlyArgLeuLysTrpLeuGlnArgLeuAlaTyrIleAsnThrIleValTyrProPheTh 760
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 TGGTCGTCTTAAATGGCTTCAAAGACTAGCATATATAAACACCATTGTCTATCCTTTCAC 2340

Qy        760 rSerLeuProLeuIleAlaTyrCysThrIleProAlaValCysLeuLeuThrGlyLysPh 780
              ||||||||||||||||||||||||||||||:::::|||||||:::|||||||||||||||||||
Db       2341 ATCCCTTCCACTCATTGCCTATTGTTCACTACCAGCAATCTGTCTTCTCACAGGAAAATT 2400

Qy        780 eIleIleProThrLeuSerAsnLeuAlaSerMetLeuPheLeuGlyLeuPheIleSerIl 800
              |||||||||||||||||||||||||||||:::|||||||||||||||||||||||:::|||||
Db       2401 TATCATACCAACGCTCTCAAACCTGGCAAGTGTTCTCTTTCTTGGCCTTTTCCTTTCCAT 2460

Qy        800 eIleValThrAlaValLeuGluLeuArgTrpSerGlyValSerIleGluAspLeuTrpAr 820
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2461 TATCGTGACTGCTGTTCTCGAGCTCCGATGGAGTGGTGTCAGCATTGAGGACTTATGGCG 2520

Qy        820 gAsnGluGlnPheTrpValIleGlyGlyValSerAlaHisLeuPheAlaValPheGlnGl 840
              ||||||||||||||||||||||||| |||||  |||||||||||||||||||||||||||||
Db       2521 TAACGAGCAGTTTTGGGTCATCGGTGGCGTTTCAGCCCATCTCTTTGCCGTCTTCCAAGG 2580

Qy        840 yPheLeuLysMetLeuAlaGlyIleAspThrAsnPheThrValThrAlaLysAlaAlaGl 860
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2581 TTTCCTTAAGATGCTTGCGGGCATTGACATCAACTTTACTGTCACTGCCAAAGCAGCTGA 2640

Qy        860 uAspAlaGluPheGlyGluLeuTyrMetValLysTrpThrSerLeuLeuIleProProTh 880
              :|||||||:::|||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2641 TGATGCAGATTTTGGTGAGCTCTACATTGTGAAATGGACTAGCTTCTTAATCCCTCCAAC 2700

Qy        880 rThrLeuLeuIleIleAsnMetSerGly  CysAlaGlyPheSerAspAlaLeuAsnLy 899
              ||||||||||||||:::||||||    ||      ||||||| |||| |||||||||||||
Db       2701 AACACTCCTCATCGTCAACATGGTTGGTGTCGTTGCCGGATCTCCGATGCCCTCAACAA 2760

Qy        899 sGlyTyrGluAlaTrpGlyPr  ||||PheLysGlyLysValIleL  eAlaPheTrpValIleLe 919
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2761 AGGGTACGAAGCTTGGGGATCC  CTTCTTAAAGTTCC  TTCCTCTGTGGTCATCCT 2820

Qy        919 uHisLeuTyrProPheLeuLysGlyLeuMetGlyArgGlnLeuTrpProThrIleVa 939
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2821 CCATCTTATCCATTCCTCAAAGGTCTTATGGGACGCCAAATACAGGACACCAACCATTGT 2880

Qy        939 lValLeuTrpSerValLeuLeuAlaSerValPheSerLeuValTrpValLysIleAsnPr 959
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2881 TGTCCTTTGGTCAGTGTTGTTGGGTTTCTGTCTTCTCTCTTGTTTGGGTTCGGATCAACCC 2940

Qy        959 oPheValAsnLysValAspAsnThrLeuValAlaGluThrCysIleSerIleAspCys 978
              |||||||:::      ||||:::|||||||||||||||||||||||||||||||||||||
Db       2941 GTTTGTCAGCACCGCCGATAGCACCACCGTGTCACAGAGCTCATTTCCATTGATTGT 2998

RESULT 3
AAV34432
ID   AAV34432 standard; cDNA; 3328 BP.
XX
AC   AAV34432;
XX
DT   28-SEP 1998  (first entry)
XX
DE   Cotton cellulose synthase CelA1 cDNA.
XX
KW   Cellulose synthase; CelA1 gene; cotton; forest tree; plant fibre;
KW   wood; transgenic plant; ds.
XX
OS   Gossypium hirsutum Acala SJ-2.
XX
FH   Key             Location/Qualifiers
FT   CDS             179..3103
FT                   /*tag=  a
XX
PN   WO9818949-A2.
XX
PD   07-MAY-1998.
XX
PF   29-OCT-1997;   97WO-US19529.
XX
PR   29-OCT-1996;   96US-0029987.
XX
PA   (CALJ ) CALGENE INC.
XX
PI   Delmer DP,  Pear JR,  Stalker DM;
XX
DR   WPI; 1998-272238/24.
DR   P-PSDB; AAW60039.
XX
PT   New isolated plant cellulose synthesis enzyme DNA - is used to
PT   develop products for producing modified plant fibre products,
PT   particularly for modifying cotton fibre and wood quality
XX
PS   Claim 5; Fig 6A-E; 49pp; English.
XX
CC   This full-length cDNA sequence corresponds to the cotton CelA1 gene
CC   transcript and encodes a 973-amino acid cellulose synthase (see
CC   AAW60039).  2 cDNA clones that exhibited small blocks of amino acid
CC   homology to proteins encoded by bacterial CelA genes were isolated
CC   from an unamplified cDNA library prepared from mRNA of fibres of
CC   cotton Acala SJ-2 harvested at 21 dpa (time of maximal secondary
CC   wall cellulose synthesis).  These 2 clones were partially homologous
CC   at the nucleotide and amino acid levels and were designated CelA1
CC   and CelA2 (see AAV34433), respectively.  The 2 cotton CelA genes are
CC   both expressed in developing fibre at approximately 500 times their
CC   level of expression in other cotton tissues, and constitute
CC   approximately 1-2% of the 24 dpa fibre mRNA.  The CelA1 gene
CC   promoter region (see AAV34434) is also claimed.  The isolated cDNA
CC   clones are used in claimed methods for modifying fibre phenotype
CC   in a cotton plant and for modifying wood quality phenotype in a
CC   forest tree species.  Lignin can be reduced by over-expressing the
CC   cellulose synthase product and diverting production into cellulose.
XX
SQ   Sequence 3328 BP; 914 A; 690 C; 763 G; 961 T; 0 other;

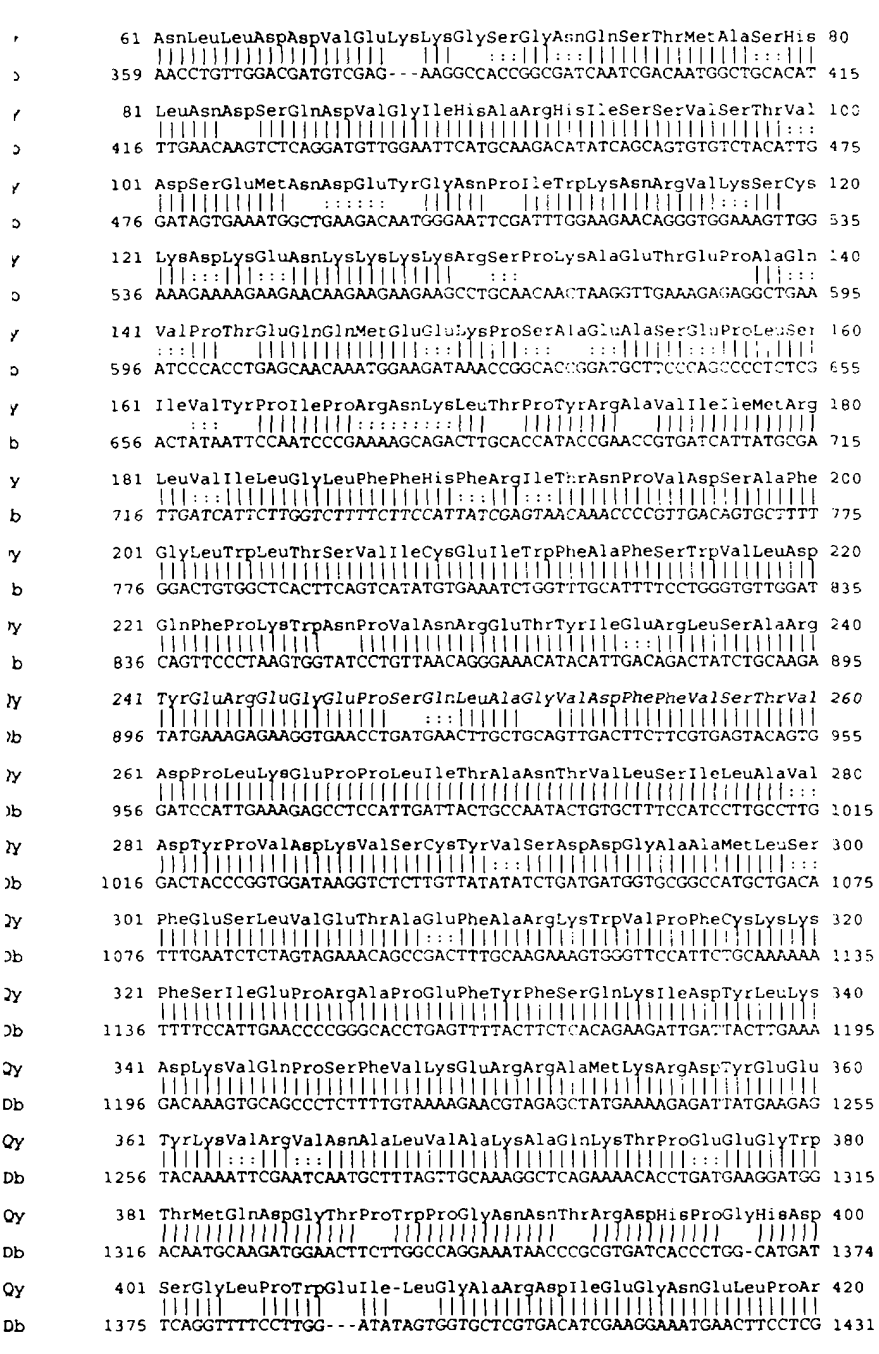
Alignment Scores:
Pred. No.:              0              Length:          3328
Score:                  4540.00        Matches:         845
Percent Similarity:     93.27%         Conservative:    69
Best Local Similarity:  86.22%         Mismatches:      59
Query Match:            86.64%         Indels:          8
DB:                     19             Gaps:            6

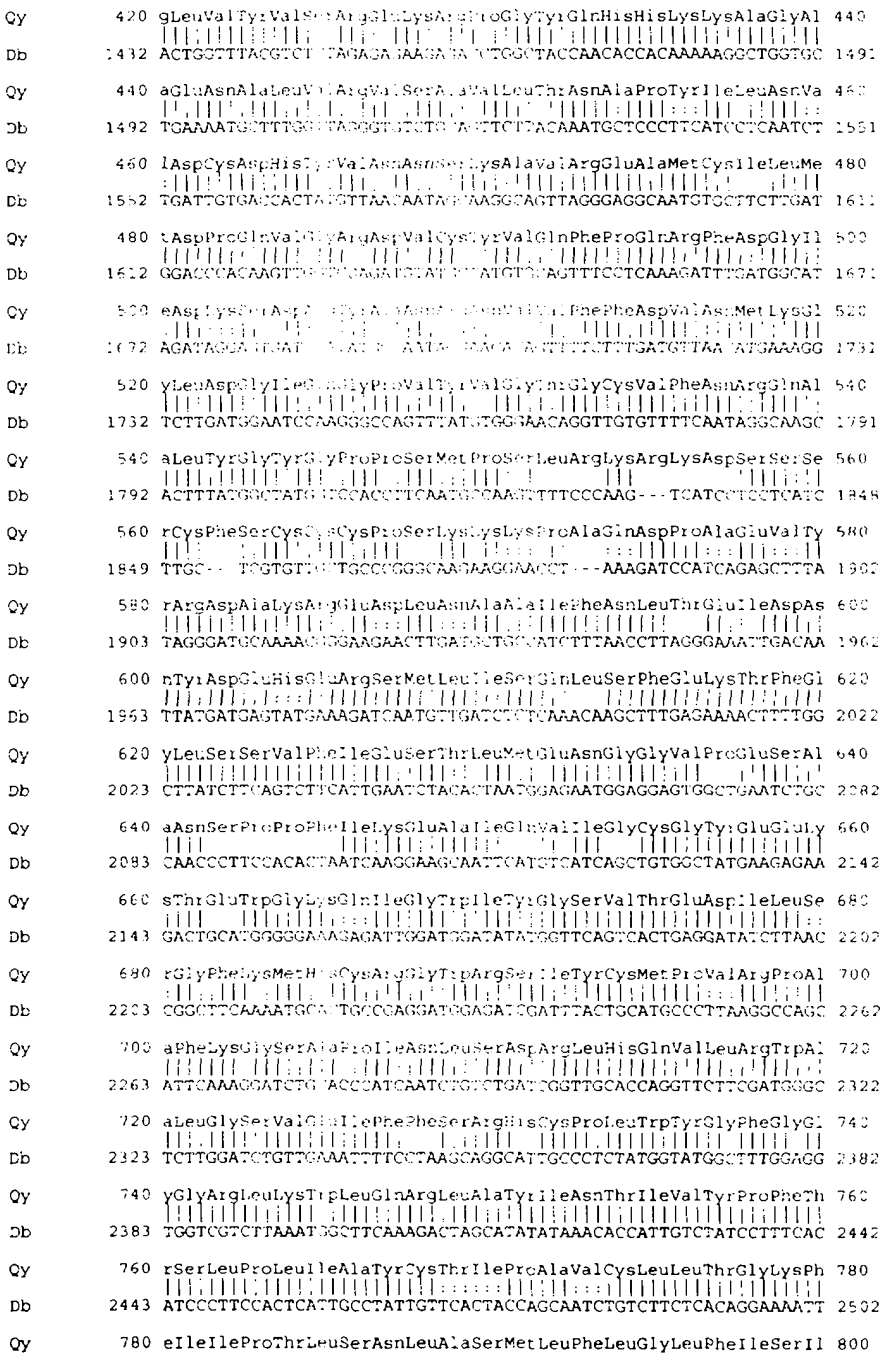
US-09-980-043-2 (1-978) x AAV34432 (1-3328)

Qy          1 MetMetGluSerGlyAlaProIleCysHisThrCysGlyGluGlnValGlyHisAspAla 20
              ||||||||||||||||   |||:::|||||||||||||||||||   ||||||   :::
Db        179 ATGATGGAATCTGGGGTTCCTGTTTGCCACACTTGTGGTGAACATGTTGGGTTGAATGTT 238

Qy         21 AsnGlyGluLeuPheValAlaCysHisGluCysSerTyrProMetCysLysSerCysPhe 40
              ||||||||   |||||||||||||||||||||||:::::|||:::|||||||||||||||
Db        239 AATGGTGAACCTTTTGTGGCTTGCCATGAATGTAATTTCCCTATTTGTAAGAGTTGTTTT 298

Qy         41 GluPheGluIleAsnGluGlyArgLysValCysLeuArgCysGlySerProTyrAspGlu 60
              |||:::::::::   ||||||||||||   |||||||||||||||||||||||||||||||
Db        299 GAGTATGATCTTAAGGAAGGACGAAAAGCTTGCTTGCGTTGTGGTAGTCCATATGATGAA 358
```


61 AsnLeuLeuAspAspValGluLysLysGlySerGlyAsnGlnSerThrMetAlaSerHis 80
359 AACCTGTTGGACGATGTCGAG---AAGGCCACCGGCGATCAATCGACAATGGCTGCACAT 415
81 LeuAsnAspSerGlnAspValGlyIleHisAlaArgHisIleSerSerValSerThrVal 100
416 TTGAACAAGTCTCAGGATGTTGGAATTCATGCAAGACATATCAGCAGTGTGTCTACATTG 475
101 AspSerGluMetAsnAspGluTyrGlyAsnProIleTrpLysAsnArgValLysSerCys 120
476 GATAGTGAAATGGCTGAAGACAATGGGAATTCGATTTGGAAGAACAGGGTGGAAAGTTGG 535
121 LysAspLysGluAsnLysLysLysLysArgSerProLysAlaGluThrGluProAlaGln 140
536 AAAGAAAAGAAGAACAAGAAGAAGAAGCCTGCAACAACTAAGGTTGAAAGAGAGGCTGAA 595
141 ValProThrGluGlnGlnMetGluGluLysProSerAlaGluAlaSerGluProLeuSer 160
596 ATCCCACCTGAGCAACAAATGGAAGATAAACCGGCACCGGATGCTTCCCAGCCCCTCTCG 655
161 IleValTyrProIleProArgAsnLysLeuThrProTyrArgAlaValIleIleMetArg 180
656 ACTATAATTCCAATCCCGAAAAGCAGACTTGCACCATACCGAACCGTGATCATTATGCGA 715
181 LeuValIleLeuGlyLeuPhePheHisPheArgIleThrAsnProValAspSerAlaPhe 200
716 TTGATCATTCTTGGTCTTTTCTTCCATTATCGAGTAACAAACCCCGTTGACAGTGCTTTT 775
201 GlyLeuTrpLeuThrSerValIleCysGluIleTrpPheAlaPheSerTrpValLeuAsp 220
776 GGACTGTGGCTCACTTCAGTCATATGTGAAATCTGGTTTGCATTTTCCTGGGTGTTGGAT 835
221 GlnPheProLysTrpAsnProValAsnArgGluThrTyrIleGluArgLeuSerAlaArg 240
836 CAGTTCCCTAAGTGGTATCCTGTTAACAGGGAAACATACATTGACAGACTATCTGCAAGA 895
241 TyrGluArgGluGlyGluProSerGlnLeuAlaGlyValAspPhePheValSerThrVal 260
896 TATGAAAGAGAAGGTGAACCTGATGAACTTGCTGCAGTTGACTTCTTCGTGAGTACAGTG 955
261 AspProLeuLysGluProProLeuIleThrAlaAsnThrValLeuSerIleLeuAlaVal 280
956 GATCCATTGAAAGAGCCTCCATTGATTACTGCCAATACTGTGCTTTCCATCCTTGCCTTG 1015
281 AspTyrProValAspLysValSerCysTyrValSerAspAspGlyAlaAlaMetLeuSer 300
1016 GACTACCCGGTGGATAAGGTCTCTTGTTATATATCTGATGATGGTGCGGCCATGCTGACA 1075
301 PheGluSerLeuValGluThrAlaGluPheAlaArgLysTrpValProPheCysLysLys 320
1076 TTTGAATCTCTAGTAGAAACAGCCGACTTTGCAAGAAAGTGGGTTCCATTCTGCAAAAAA 1135
321 PheSerIleGluProArgAlaProGluPheTyrPheSerGlnLysIleAspTyrLeuLys 340
1136 TTTTCCATTGAACCCCGGGCACCTGAGTTTTACTTCTCACAGAAGATTGATTACTTGAAA 1195
341 AspLysValGlnProSerPheValLysGluArgArgAlaMetLysArgAspTyrGluGlu 360
1196 GACAAAGTGCAGCCCTCTTTTGTAAAAGAACGTAGAGCTATGAAAAGAGATTATGAAGAG 1255
361 TyrLysValArgValAsnAlaLeuValAlaLysAlaGlnLysThrProGluGluGlyTrp 380
1256 TACAAAATTCGAATCAATGCTTTAGTTGCAAAGGCTCAGAAAACACCTGATGAAGGATGG 1315
381 ThrMetGlnAspGlyThrProTrpProGlyAsnAsnThrArgAspHisProGlyHisAsp 400
1316 ACAATGCAAGATGGAACTTCTTGGCCAGGAAATAACCCGCGTGATCACCCTGG-CATGAT 1374
401 SerGlyLeuProTrpGluIle-LeuGlyAlaArgAspIleGluGlyAsnGluLeuProAr 420
1375 TCAGGTTTTCCTTGG---ATATAGTGGTGCTCGTGACATCGAAGGAAATGAACTTCCTCG 1431


Qy 420 gLeuValTyrValSerArgGluLysAspProGlyTyrGlnHisHisLysLysAlaGlyAl 440
Db 1432 ACTGGTTTACGTCTTAGAGAGAAGATCTGGCTACCAACACCACAAAAAGGCTGGTGC 1491
Qy 440 aGluAsnAlaLeuValArgValSerAlaValLeuThrAsnAlaProTyrIleLeuAsnVa 460
Db 1492 TGAAAATGCTTTGGTAGGTGTCTCTACTTCTTACAAATGCTCCCTTCATCCTCAATCT 1551
Qy 460 lAspCysAspHisTyrValAsnAsnSerLysAlaValArgGluAlaMetCysIleLeuMe 480
Db 1552 TGATTGTGACCACTATGTTAACAATAGCAAGGCAGTTAGGGAGGCAATGTGCTTCTTGAT 1611
Qy 480 tAspProGlnValGlyArgAspValCysTyrValGlnPheProGlnArgPheAspGlyIl 500
Db 1612 GGACCCACAAGTTGTCGAGATGTATCATGTCAGTTTCCTCAAAGATTTGATGGCAT 1671
Qy 500 eAspLysProAspPheGlyLeuAspMetGlyMetProValPhePheAspValAsnMetLysGl 520
Db 1672 AGATAGGATGATCAATGAACAAGTTTTCTTTGATGTTAATATGAAAGG 1731
Qy 520 yLeuAspGlyIleGlnGlyProValTyrValGlyThrGlyCysValPheAsnArgGlnAl 540
Db 1732 TCTTGATGGAATCCAAGGGCCAGTTTATGTGGGAACAGGTTGTGTTTTCAATAGGCAAGC 1791
Qy 540 aLeuTyrGlyTyrGlyProProSerMetProSerLeuArgLysArgLysAspSerSerSe 560
Db 1792 ACTTTATGGCTATGGTCCACCTTCAATGCCAAGTTTTCCCAAG---TCATCCTCCTCATC 1848
Qy 560 rCysPheSerCysGlyCysProSerLysLysLysProAlaGlnAspProAlaGluValTy 580
Db 1849 TTGC--TCGTGTTCCTGCCCGGGCAAGAAGGAACCT---AAAGATCCATCAGAGCTTTA 1902
Qy 580 rArgAspAlaLysArgGluAspLeuAsnAlaAlaIlePheAsnLeuThrGluIleAspAs 600
Db 1903 TAGGGATGCAAAAGCGGAAGAACTTGAAGCTGCCATCTTTAACCTTAGGGAAATTGACAA 1962
Qy 600 nTyrAspGluHisGluArgSerMetLeuIleSerGlnLeuSerPheGluLysThrPheGl 620
Db 1963 TTATGATGAGTATGAAAGATCAATGTTGATCTCTCAAACAAGCTTTGAGAAAACTTTTGG 2022
Qy 620 yLeuSerSerValPheIleGluSerThrLeuMetGluAsnGlyGlyValProGluSerAl 640
Db 2023 CTTATCTTCAGTCTTCATTGAATCTACACTAATGGAGAATGGAGGAGTGGCTGAATCTGC 2082
Qy 640 aAsnSerProProPheIleLysGluAlaIleGlnValIleGlyCysGlyTyrGluGluLy 660
Db 2083 CAACCCTTCCACACTAATCAAGGAAGCAATTCATGTCATCAGCTGTGGCTATGAAGAGAA 2142
Qy 660 sThrGluTrpGlyLysGlnIleGlyTrpIleTyrGlySerValThrGluAspIleLeuSe 680
Db 2143 GACTGCATGGGGGAAAGAGATTGGATGGATATATGGTTCAGTCACTGAGGATATCTTAAC 2202
Qy 680 rGlyPheLysMetHisCysArgGlyTrpArgSerIleTyrCysMetProValArgProAl 700
Db 2203 CGGCTTCAAAATGCATGCCGAGGATGGAGATCGATTTACTGCATGCCCTTAAGGCCAGC 2262
Qy 700 aPheLysGlySerAlaProIleAsnLeuSerAspArgLeuHisGlnValLeuArgTrpAl 720
Db 2263 ATTCAAAGGATCTGTACCCATCAATCTGTCTGATCGGTTGCACCAGGTTCTTCGATGGGC 2322
Qy 720 aLeuGlySerValGluIlePhePheSerArgHisCysProLeuTrpTyrGlyPheGlyGl 740
Db 2323 TCTTGGATCTGTTGAAATTTTCCTAAGCAGGCATTGCCCTCTATGGTATGGCTTTGGAGG 2382
Qy 740 yGlyArgLeuLysTrpLeuGlnArgLeuAlaTyrIleAsnThrIleValTyrProPheTh 760
Db 2383 TGGTCGTCTTAAATGGCTTCAAAGACTAGCATATATAAACACCATTGTCTATCCTTTCAC 2442
Qy 760 rSerLeuProLeuIleAlaTyrCysThrIleProAlaValCysLeuLeuThrGlyLysPh 780
Db 2443 ATCCCTTCCACTCATTGCCTATTGTTCACTACCAGCAATCTGTCTTCTCACAGGAAAATT 2502
Qy 780 eIleIleProThrLeuSerAsnLeuAlaSerMetLeuPheLeuGlyLeuPheIleSerIl 800

```
Db       2503 TATCATACCAACGCTCTCAAACCTGGCAAGTGTTCTCTTTCTTGGCCTTTTCCTTTCCAT 2562

Qy        800 eIleValThrAlaValLeuGluLeuArgTrpSerGlyValSerIleGluAspLeuTrpAr 820

Db       2563 TATCGTGACTGCTGTTCTCGAGCTCCGATGGAGTGGTGTCAGCATTGAGGACTTATGGCG 2622

Qy        820 gAsnGluGlnPheTrpValIleGlyGlyValSerAlaHisLeuPheAlaValPheGlnGl 840

Db       2623 TAACGAGCAGTTTTGGGTCATCGGTGGCGTTTCAGCCCATCTCTTTGCCGTCTTCCAAGG 2682

Qy        840 yPheLeuLysMetLeuAlaGlyIleAspThrAsnPheThrValThrAlaLysAlaAlaGl 860

Db       2683 TTTCCTTAAGATGCTTGCGGGCATTGACACCAACTTTACTGTCACTGCCAAAGCAGCTGA 2742

Qy        860 uAspAlaGluPheGlyGluLeuTyrMetValLysTrpThrThrLeuLeuIleProProTh 880

Db       2743 TGATGCAGATTTTGGTGAGCTCTACATTGTGAAATGGACTACACTTCTAATCCCTCCAAC 2802

Qy        880 rThrLeuLeuIleIleAsnMetSerGly---CysAlaGlyPheSerAspAlaLeuAsnLy 899

Db       2803 AACACTCCTCATCGTCAACATGGTTGGTGTCGTTGCCGGATTCTCCGATGCCCTCAACAA 2862

Qy        899 sGlyTyrGluAlaTrpGlyProLeuPheGlyLysValPhePheAlaPheTrpValIleLe 919

Db       2863 AGGGTACGAAGCTTGGGGACCACTCTTTGGCAAAGTGTTCTTTTCCTTCTGGGTCATCCT 2922

Qy        919 uHisLeuTyrProPheLeuLysGlyLeuMetGlyArgGlnAsnLeuThrProThrIleVa 939

Db       2923 CCATCTTTATCCATTCCTCAAAGGTCTTATGGGACGCCAAAACAGGACACCAACCATTGT 2982

Qy        939 lValLeuTrpSerValLeuLeuAlaSerValPheSerLeuValTrpValLysIleAsnPr 959

Db       2983 TGTCCTTTGGTCAGTGTTGTTGGCTTCTGTCTTCTCTCTTGTTTGGGTTCGGATCAACCC 3042

Qy        959 oPheValAsnLysValAspAsnThrLeuValAlaGluThrCysIleSerIleAspCys 978

Db       3043 GTTTGTCAGCACCGCCGATAGCACCACCGTGTCACAGAGCTGCATTTCCATTGATTGT 3100
```

RESULT 4

AAA67145

```
ID   AAA67145 standard; DNA; 3747 BP.
XX
AC   AAA67145;
XX
DT   31-OCT 2000  (first entry)
XX
DE   Eucalyptus grandis cellulose synthase nucleotide sequence SEQ ID NO:141.
XX
KW   Eucalyptus grandis; pinus radiata; Monterey pine; modification;
KW   plant cell wall; polysaccharide; polysaccharide biosynthetic pathway;
KW   transgenic plant; ds.
XX
OS   Eucalyptus grandis.
XX
PN   WO200022092-A2.
XX
PD   20 APR 2000.
XX
PF   08-OCT-1999;   99WO-NZ00169.
XX
PR   13-OCT-1998;   98US-0170862.
PR   11-AUG-1999;   99US-0148426.
XX
PA   (GENE-) GENESIS RES & DEV CORP LTD.
PA   (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI   Bloksberg LN;
XX
DR   WPI; 2000-339328/29.
DR   P-PSDB; AAB16338.
XX
PT   New genes encoding proteins involved in a plant polysaccharide
PT   biosynthetic pathway, useful for modulating or altering the
PT   polysaccharide content, composition or structure of the plant -
XX
PS   Claim 1; Page 106-107; 301pp; English.
XX
CC   The present invention describes isolated polynucleotides (PN) comprising
CC   a sequence selected from one of 835 nucleotide sequences given in
CC   AAA67073 to AAA67907, their (reverse) complements, sequences producing
CC   an Expectation (E) value of 0.01 or less compared to the 835 sequences,
CC   sequences at least 50% identical to them, 200, 100, 40 or 20-mers of the
CC   835 sequences or sequences that are degenerately equivalent or allelic
CC   to the 835 sequences. The polynucleotides are used to modify the
CC   activity of a polypeptide involved in a polysaccharide biosynthetic
CC   pathway in the plant. They are especially used to modulate or alter the
CC   polysaccharide content, composition or structure of the plant. AAB16268
CC   to AAB16340 are proteins encoded by some of the polynucleotide sequence
CC   given in the present invention.
XX
SQ   Sequence 3747 BP; 976 A; 822 C; 935 G; 1014 T; 0 other;
```

Alignment Scores:

Pred. No.:	0	Length:	3747
Score:	4368.00	Matches:	822
Percent Similarity:	91.45%	Conservative:	76
Best Local Similarity:	83.71%	Mismatches:	77
Query Match:	83.36%	Indels:	11
DB:	21	Gaps:	5

US-09-980-043-2 (1-978) x AAA67145 (1-3747)

```
Qy          1 MetMetGluSerGlyAlaProIleCysHisThrCysGlyGluGlnValGlyHisAspAla 20

Db        511 ATGATGGAATCCGGGGTTCCCCTGTGCAACACTTGCGGAGAGGCTGTTGGGGTTGATGAG 570

Qy         21 AsnGlyGluLeuPheValAlaCysHisGluCysSerTyrProMetCysLysSerCysPhe 40

Db        571 AAAGGCGAGGTCTTCGTGGCTTGTCAAGAGTGCAACTTCGCCATTTGCAAGGCTTGTGTC 630

Qy         41 GluPheGluIleAsnGluGlyArgLysValCysLeuArgCysGlySerProTyrAspGlu 60

Db        631 GAATATGAGATTAAGGAAGGAAGAAAAGCGTGCTTGCGCTGTGGCACTCCATTTGAAGCG 690

Qy         61 AsnLeuLeuAspAspValGluLysLysGlySerGlyAsnGlnSerThrMetAlaSerHis 80

Db        691 AACTCGATGGCTGATGCTGAGAGAAATGAATTGGGAAGTCGATCGACAATGGCAGCTCAA 750

Qy         81 LeuAsnAspSerGlnAspValGlyIleHisAlaArgHisIleSerSerValSerThrVal 100

Db        751 CTCAATGATCCTCAGGACACAGGGATTCATGCTAGACACATCAGCAGTGTTTCTACGTTG 810

Qy        101 AspSerGluMetAsnAspGluTyrGlyAsnProIleTrpLysAsnArgValLysSerCys 120

Db        811 GATAGTGAATACAATGATGAGACTGGGAACCCTATCTGGAAGAATAGAGTGGAGAGCTGG 870

Qy        121 LysAspLysGluAsnLysLysLysLysArgSerProLysAlaGluThrGluProAlaGln 140

Db        871 AAGGACAAAAAGAATAAGAAGAAGAAGGCCCCGACGAAGGCTGAGAAAGAG---GCTCAA 927

Qy        141 ValProThrGluGlnGlnMetGluGluLysProSerAlaGluAlaSerGluProLeuSer 160

Db        928 GTTCCACCAGAGCAGCAGATGGAAGAGAAGCAAATTGCTGATGCTTCAGAGCCACTCTCG 987

Qy        161 IleValTyrProIleProArgAsnLysLeuThrProTyrArgAlaValIleIleMetArg 180

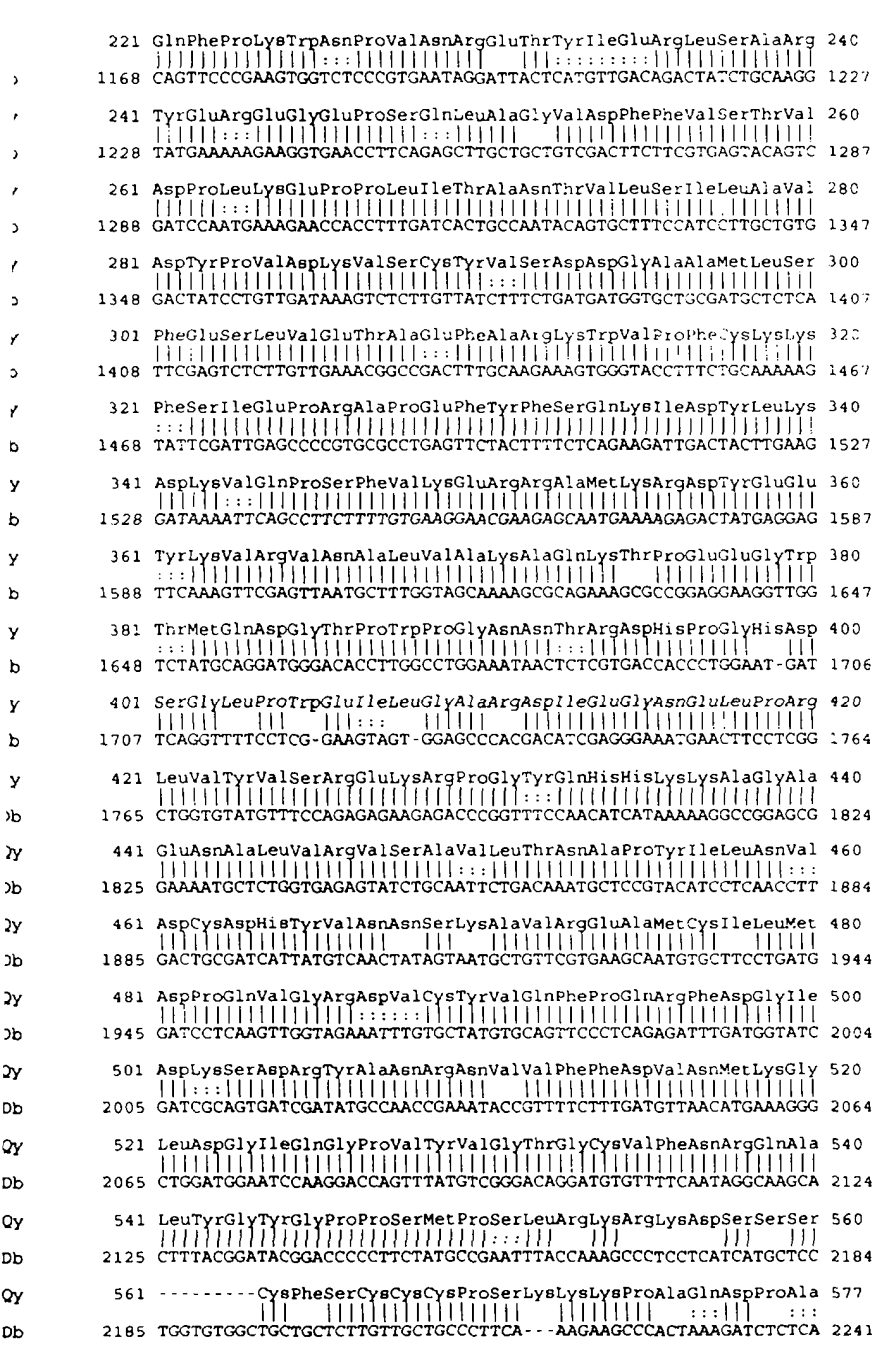
Db        988 ACCGTTATTCCCATTGCCAAAAGCAAACTCGCACCATACCGAACCGTAATAATAATGCGA 1047

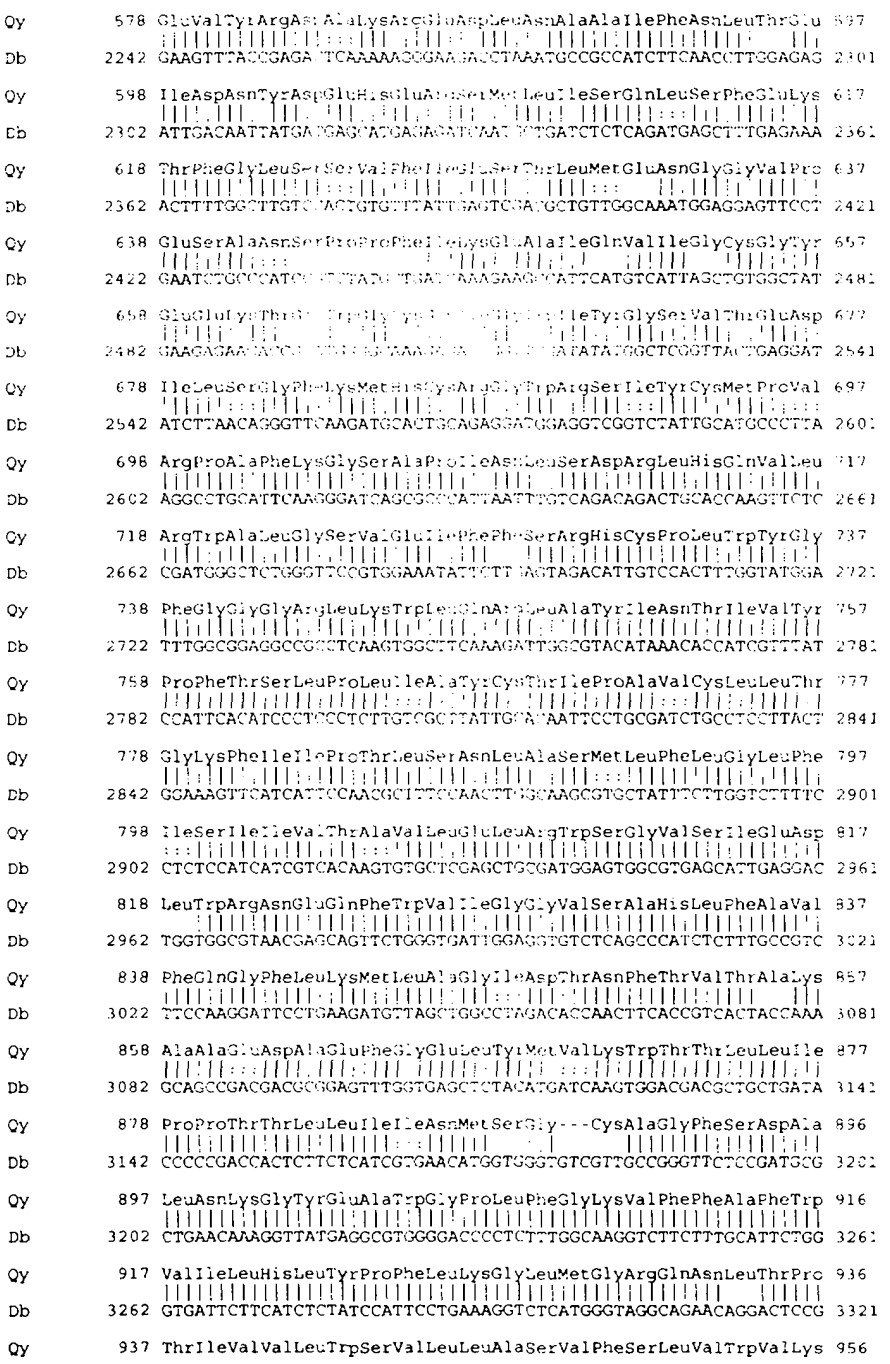
Qy        181 LeuValIleLeuGlyLeuPhePheHisPheArgIleThrAsnProValAspSerAlaPhe 200

Db       1048 TTGATCATTTTGGCACTTTTCTTCCATTACCGAGTGACACATCCAGTCGACAGCGCCTAT 1107

Qy        201 GlyLeuTrpLeuThrSerValIleCysGluIleTrpPheAlaPheSerTrpValLeuAsp 220

Db       1108 CCATTGTGGCTTACATCCATCATATGTGAGATTTGGTTCGCATACTCATGGGTGTTGGAT 1167
```


221 GlnPheProLysTrpAsnProValAsnArgGluThrTyrIleGluArgLeuSerAlaArg 240
1168 CAGTTCCCGAAGTGGTCTCCCGTGAATAGGATTACTCATGTTGACAGACTATCTGCAAGG 1227
241 TyrGluArgGluGlyGluProSerGlnLeuAlaGlyValAspPhePheValSerThrVal 260
1228 TATGAAAAAGAAGGTGAACCTTCAGAGCTTGCTGCTGTCGACTTCTTCGTGAGTACAGTC 1287
261 AspProLeuLysGluProProLeuIleThrAlaAsnThrValLeuSerIleLeuAlaVal 280
1288 GATCCAATGAAAGAACCACCTTTGATCACTGCCAATACAGTGCTTTCCATCCTTGCTGTG 1347
281 AspTyrProValAspLysValSerCysTyrValSerAspAspGlyAlaAlaMetLeuSer 300
1348 GACTATCCTGTTGATAAAGTCTCTTGTTATCTTTCTGATGATGGTGCTGCGATGCTCTCA 1407
301 PheGluSerLeuValGluThrAlaGluPheAlaArgLysTrpValProPheCysLysLys 320
1408 TTCGAGTCTCTTGTTGAAACGGCCGACTTTGCAAGAAAGTGGGTACCTTTCTGCAAAAAG 1467
321 PheSerIleGluProArgAlaProGluPheTyrPheSerGlnLysIleAspTyrLeuLys 340
1468 TATTCGATTGAGCCCCGTGCGCCTGAGTTCTACTTTTCTCAGAAGATTGACTACTTGAAG 1527
341 AspLysValGlnProSerPheValLysGluArgArgAlaMetLysArgAspTyrGluGlu 360
1528 GATAAAATTCAGCCTTCTTTTGTGAAGGAACGAAGAGCAATGAAAAGAGACTATGAGGAG 1587
361 TyrLysValArgValAsnAlaLeuValAlaLysAlaGlnLysThrProGluGluGlyTrp 380
1588 TTCAAAGTTCGAGTTAATGCTTTGGTAGCAAAAGCGCAGAAAGCGCCGGAGGAAGGTTGG 1647
381 ThrMetGlnAspGlyThrProTrpProGlyAsnAsnThrArgAspHisProGlyHisAsp 400
1648 TCTATGCAGGATGGGACACCTTGGCCTGGAAATAACTCTCGTGACCACCCTGGAAT-GAT 1706
401 SerGlyLeuProTrpGluIleLeuGlyAlaArgAspIleGluGlyAsnGluLeuProArg 420
1707 TCAGGTTTTCCTCG-GAAGTAGT-GGAGCCCACGACATCGAGGGAAATGAACTTCCTCGG 1764
421 LeuValTyrValSerArgGluLysArgProGlyTyrGlnHisHisLysLysAlaGlyAla 440
1765 CTGGTGTATGTTTCCAGAGAGAAGAGACCCGGTTTCCAACATCATAAAAAGGCCGGAGCG 1824
441 GluAsnAlaLeuValArgValSerAlaValLeuThrAsnAlaProTyrIleLeuAsnVal 460
1825 GAAAATGCTCTGGTGAGAGTATCTGCAATTCTGACAAATGCTCCGTACATCCTCAACCTT 1884
461 AspCysAspHisTyrValAsnAsnSerLysAlaValArgGluAlaMetCysIleLeuMet 480
1885 GACTGCGATCATTATGTCAACTATAGTAATGCTGTTCGTGAAGCAATGTGCTTCCTGATG 1944
481 AspProGlnValGlyArgAspValCysTyrValGlnPheProGlnArgPheAspGlyIle 500
1945 GATCCTCAAGTTGGTAGAAATTTGTGCTATGTGCAGTTCCCTCAGAGATTTGATGGTATC 2004
501 AspLysSerAspArgTyrAlaAsnArgAsnValValPhePheAspValAsnMetLysGly 520
2005 GATCGCAGTGATCGATATGCCAACCGAAATACCGTTTTCTTTGATGTTAACATGAAAGGG 2064
521 LeuAspGlyIleGlnGlyProValTyrValGlyThrGlyCysValPheAsnArgGlnAla 540
2065 CTGGATGGAATCCAAGGACCAGTTTATGTCGGGACAGGATGTGTTTTCAATAGGCAAGCA 2124
541 LeuTyrGlyTyrGlyProProSerMetProSerLeuArgLysArgLysAspSerSerSer 560
2125 CTTTACGGATACGGACCCCCTTCTATGCCGAATTTACCAAAGCCCTCCTCATCATGCTCC 2184
561 ---------CysPheSerCysCysCysProSerLysLysLysProAlaGlnAspProAla 577
2185 TGGTGTGGCTGCTGCTCTTGTTGCTGCCCTTCA---AAGAAGCCCACTAAAGATCTCTCA 2241


Qy 578 GluValTyrArgAsnAlaLysArgGluAspLeuAsnAlaAlaIlePheAsnLeuThrGlu 597
Db 2242 GAAGTTTACCGAGATTCAAAAAGGGAAGACCTAAATGCCGCCATCTTCAACCTTGGAGAG 2301
Qy 598 IleAspAsnTyrAspGluHisGluAsnSerMetLeuIleSerGlnLeuSerPheGluLys 617
Db 2302 ATTGACAATTATGATGAGCATGAGAATCCAACTGATCTCTCAGATGAGCTTTGAGAAA 2361
Qy 618 ThrPheGlyLeuSerSerValPheLeuGluSerThrLeuMetGluAsnGlyValPro 637
Db 2362 ACTTTTGGCTTGTCAAGTGTTTATTGAGTCGAAGCTGTTGGCAAATGGAGGAGTTCCT 2421
Qy 638 GluSerAlaAsnSerProProPheLeuLysGluAlaIleGlnValIleGlyCysGlyTyr 657
Db 2422 GAATCTGCCCATCCTTATCTGAAAGAAGCCATTCATGTCATTAGCTGTGGCTAT 2481
Qy 658 GluGluLysThrSerTrpGlyLysGlyIleTyrGlySerValThrGluAsp 677
Db 2482 GAAGAGAAAACCTCCTGGGAAAAGGGAATATATGGCTCCGTTACTGAGGAT 2541
Qy 678 IleLeuSerGlyPheLysMetHisCysArgGlyTrpArgSerIleTyrCysMetProVal 697
Db 2542 ATCTTAACAGGGTTCAAGATGCACTGCAGAGGATGGAGGTCGGTCTATTGCATGCCCTTA 2601
Qy 698 ArgProAlaPheLysGlySerAlaProIleAsnLeuSerAspArgLeuHisGlnValLeu 717
Db 2602 AGGCCTGCATTCAAGGGATCAGCGCCCATTAATTTGTCAGACAGACTGCACCAAGTTCTC 2661
Qy 718 ArgTrpAlaLeuGlySerValGluIlePhePheSerArgHisCysProLeuTrpTyrGly 737
Db 2662 CGATGGGCTCTGGGTTCCGTGGAAATATTCTTTAGTAGACATTGTCCACTTTGGTATGGA 2721
Qy 738 PheGlyGlyGlyArgLeuLysTrpLeuGlnArgLeuAlaTyrIleAsnThrIleValTyr 757
Db 2722 TTTGGCGGAGGCCGCCTCAAGTGGCTTCAAAGATTGGCGTACATAAACACCATCGTTTAT 2781
Qy 758 ProPheThrSerLeuProLeuIleAlaTyrCysThrIleProAlaValCysLeuLeuThr 777
Db 2782 CCATTCACATCCCTCCCTCTTGTCGCTTATTGTATAATTCCTGCGATCTGCCTCCTTACT 2841
Qy 778 GlyLysPheIleIleProThrLeuSerAsnLeuAlaSerMetLeuPheLeuGlyLeuPhe 797
Db 2842 GGAAAGTTCATCATTCCAACGCTTTCCAACTTTGCAAGCGTGCTATTTCTTGGTCTTTTC 2901
Qy 798 IleSerIleIleValThrAlaValLeuGluLeuArgTrpSerGlyValSerIleGluAsp 817
Db 2902 CTCTCCATCATCGTCACAAGTGTGCTCGAGCTGCGATGGAGTGGCGTGAGCATTGAGGAC 2961
Qy 818 LeuTrpArgAsnGluGlnPheTrpValIleGlyGlyValSerAlaHisLeuPheAlaVal 837
Db 2962 TGGTGGCGTAACGAGCAGTTCTGGGTGATTGGAGGTGTCTCAGCCCATCTCTTTGCCGTC 3021
Qy 838 PheGlnGlyPheLeuLysMetLeuAlaGlyIleAspThrAsnPheThrValThrAlaLys 857
Db 3022 TTCCAAGGATTCCTGAAGATGTTAGCTGGCCTAGACACCAACTTCACCGTCACTACCAAA 3081
Qy 858 AlaAlaGluAspAlaGluPheGlyGluLeuTyrMetValLysTrpThrThrLeuLeuIle 877
Db 3082 GCAGCCGACGACGCGGAGTTTGGTGAGCTCTACATGATCAAGTGGACGACGCTGCTGATA 3141
Qy 878 ProProThrThrLeuLeuIleIleAsnMetSerGly---CysAlaGlyPheSerAspAla 896
Db 3142 CCCCCGACCACTCTTCTCATCGTGAACATGGTGGGTGTCGTTGCCGGGTTCTCCGATGCG 3201
Qy 897 LeuAsnLysGlyTyrGluAlaTrpGlyProLeuPheGlyLysValPhePheAlaPheTrp 916
Db 3202 CTGAACAAAGGTTATGAGGCGTGGGGACCCCTCTTTGGCAAGGTCTTCTTTGCATTCTGG 3261
Qy 917 ValIleLeuHisLeuTyrProPheLeuLysGlyLeuMetGlyArgGlnAsnLeuThrPro 936
Db 3262 GTGATTCTTCATCTCTATCCATTCCTGAAAGGTCTCATGGGTAGGCAGAACAGGACTCCG 3321
Qy 937 ThrIleValValLeuTrpSerValLeuLeuAlaSerValPheSerLeuValTrpValLys 956

```
Db       3322 ACCATTGTGGTTCTCTGGTCGGTGTT-CTGGCTTCTGTCTTCTCTCTCGTCTGGGTGAAG 3380

Qy        957 IleAsnProPheValAsnLysValAspAsnThrLeuValAlaGluThrCysIleSerIle 976

Db       3381 ATCGATCCATTCGTGAGCAAATCCGATGCTGACCTC---TCCCAGAGCTGCAGTTCTATA 3437

Qy        977 AspCys 978

Db       3438 GATTGT 3443

RESULT 5
AAV08373
ID   AAV08373 standard; cDNA to mRNA; 3311 BP.
XX
AC   AAV08373;
XX
DT   08-FEB-1999  (first entry)
XX
DE   Cellulose synthase PcsA2 coding sequence.
XX
KW   Cellulose synthase; cotton; ds.
XX
OS   Gossypium hirsutum.
XX
FH   Key             Location/Qualifiers
FT   CDS             22..3142
FT                   /*tag=  a
XX
PN   EP875575-A2.
XX
PD   04-NOV-1998.
XX
PF   31-MAR-1998;   98EP-0302489.
XX
PR   01-APR-1997;   97JP-0083133.
XX
PA   (NISN ) NISSHINBO IND INC.
XX
PI   Aotsuka S,  Hasegawa O,  Hayashi T,  Imura Y;
XX
DR   WPI; 1998-559443/48.
DR   P-PSDB; AAW73309.
XX
PT   New DNA encoding cotton cellulose synthase proteins - useful for
PT   controlling cellulose synthesis in prokaryotic or eukaryotic cells
XX
PS   Example 5; Page 22-27; 51pp; English
XX
CC   This sequence encodes a cellulose synthase of the invention. The DNA can
CC   be used for controlling cellulose synthesis in a cell by introducing the
CC   DNA into the cell, expressing RNA with a nucleotide sequence homologous
CC   to the DNA or a nucleotide sequence complementary to the DNA. The cells
CC   are prokaryotic cells i.e. acetobacterium and/or eukaryotic cells such as
CC   yeasts i.e. Saccharomyces spp., plant cells (cotton plants) and cultured
CC   cells of mammals etc.
XX
SQ   Sequence 3311 BP; 900 A; 678 C; 788 G; 945 T; 0 other;

Alignment Scores:
Pred. No.:              0             Length:          3311
Score:                  3492.50       Matches:         1168
Percent Similarity:     77.00%        Conservative:     182
Best Local Similarity:  65.11%        Mismatches:       176
Query Match:            66.65%        Indels:            65
DB:                     19            Gaps:              18

US-09-980-043-2 (1-978) x AAV08373 (1-3311)

Qy          3 GluSerGlyAlaProIleCysHisThrCysGlyGluGlnValGlyHisAspAlaAsnGly 22

Db         95 CAATCATCAACGAAAATATGCAGGGTGTGTGGGGATAAGATGGGCAAAAGGAAAACGGA 154

Qy         23 GluLeuPheValAlaCysHisGluCysSerTyrProMetCysLysSerCysPheGluPhe 42

Db        155 CAACCGTTCGTGGCTTGTCATGTCTGTGCTTTCCCGGTTTGCCGTCCTTGTTATGAATAT 214

Qy         43 GluIleAsnGluGlyArgLysValCysLeuArgCysGlySerProTyr------------ 58

Db        215 GAAAGGAGTGAAGGAAACCAGTGCTGTCCTCAGTGCAATACTCGCTATAAGCGTCACAAA 274

Qy         59 ----------------------------------------AspGluAsnLeuLeuAspAsp 65

Db        275 GGTAGTCCAAGAATTTCAGGAGATGAAGAAGATGATTCAGATCAAGATGATTTTGATGAT 334

Qy         66 ValGluLysLysGlySerGlyAsnGlnSerThrMetAlaSerHisLeuAsnAspSerGln 85

Db        335 GAATTTCAGATTAAGAACCGCAAGGATGACTCCCATCCACAACATGAAAATGAGGAATAT 394

Qy         86 AspValGlyIleHisAlaArgHis----------------------IleSerSerVal 97

Db        395 AATAATAATAATCATCAATGGCATCCCAATGGTCAAGCTTTCTCAGTTGCCGGAAGCACG 454

Qy         98 SerThrValAspSerGluMetAsnAspGlu---TyrGlyAsnProIleTrpLysAsnArg 116

Db        455 GCGGGGAAGGATTTGGAAGGGGATAAAGAGATTTACGGAAGCGAAGAATGGAAAGAAAGA 514

Qy        117 ValLysSerCysLysAspLysGluAsnLysLysLysLysArgSerProLysAlaGluThr 136

Db        515 GTTGAGAAATGGAAAGTCAGGCAA---------GAAAAAAGAGGTTTGGTAAGCAACGAT 565

Qy        137 GluProAlaGlnValProThrGluGlnGlnMetGluGluLysProSerAlaGluAlaSer 156

Db        566 AATGGCGGAAATGATCCTCCTGAAGAA------GATGATTATCTCTTGGCTGAAGCTCGC 619

Qy        157 GluProLeuSerIleValTyrProIleProArgAsnLysLeuThrProTyrArgAlaVal 176

Db        620 CAGCCTCTATGGCGAAAAGTGCCAATTTCGTCAAGTCTGATAAGCCCTTACCGGATAGTC 679

Qy        177 IleIleMetArgLeuValIleLeuGlyLeuPhePheHisPheArgIleThrAsnProVal 196

Db        680 ATCGTCCTCCGATTCTTCATCCTCGCATTTTTCCTCCGGTTCCGTATTCTAACACCCGCC 739

Qy        197 AspSerAlaPheGlyLeuTrpLeuThrSerValIleCysGluIleTrpPheAlaPheSer 216

Db        740 TACGACGCTTACCCGTTATGGCTAATCTCTGTCATCTGCGAAGTTTGGTTCGCCTTCTCC 799

Qy        217 TrpValLeuAspGlnPheProLysTrpAsnProValAsnArgGluThrTyrIleGluArg 236

Db        800 TGGATTCTCGATCAGTTCCCTAAATGGTTCCCTATTACTCGCGAAACTTACCTCGATCGC 859

Qy        237 LeuSerAlaArgTyrGluArgGluGlyGluProSerGlnLeuAlaGlyValAspPhePhe 256

Db        860 CTCTCCTTGAGGTTCGAACGTGAAGGAGAGCCCAATCAACTTGGCCCCGTCGACGTCTTC 919

Qy        257 ValSerThrValAspProLeuLysGluProProLeuIleThrAlaAsnThrValLeuSer 276

Db        920 GTCAGTACCGTTGACCTTCTCAAGGAACCCCCCATCATAACCGCCAACGCGGTTCTATCG 979

Qy        277 IleLeuAlaValAspTyrProValAspLysValSerCysTyrValSerAspAspGlyAla 296

Db        980 ATCTTGGCCGTCGATTACCCGGTCGAGAAAGTGTGTTGTTATGTGTCGGACGATGGTGCT 1039

Qy        297 AlaMetLeuSerPheGluSerLeuValGluThrAlaGluPheAlaArgLysTrpValPro 316

Db       1040 TCCATGCTTCTTTTCGATTCGTTGTCTGAAACGGCTGAGTTCGCGAGGAGATGGGTTCCG 1099

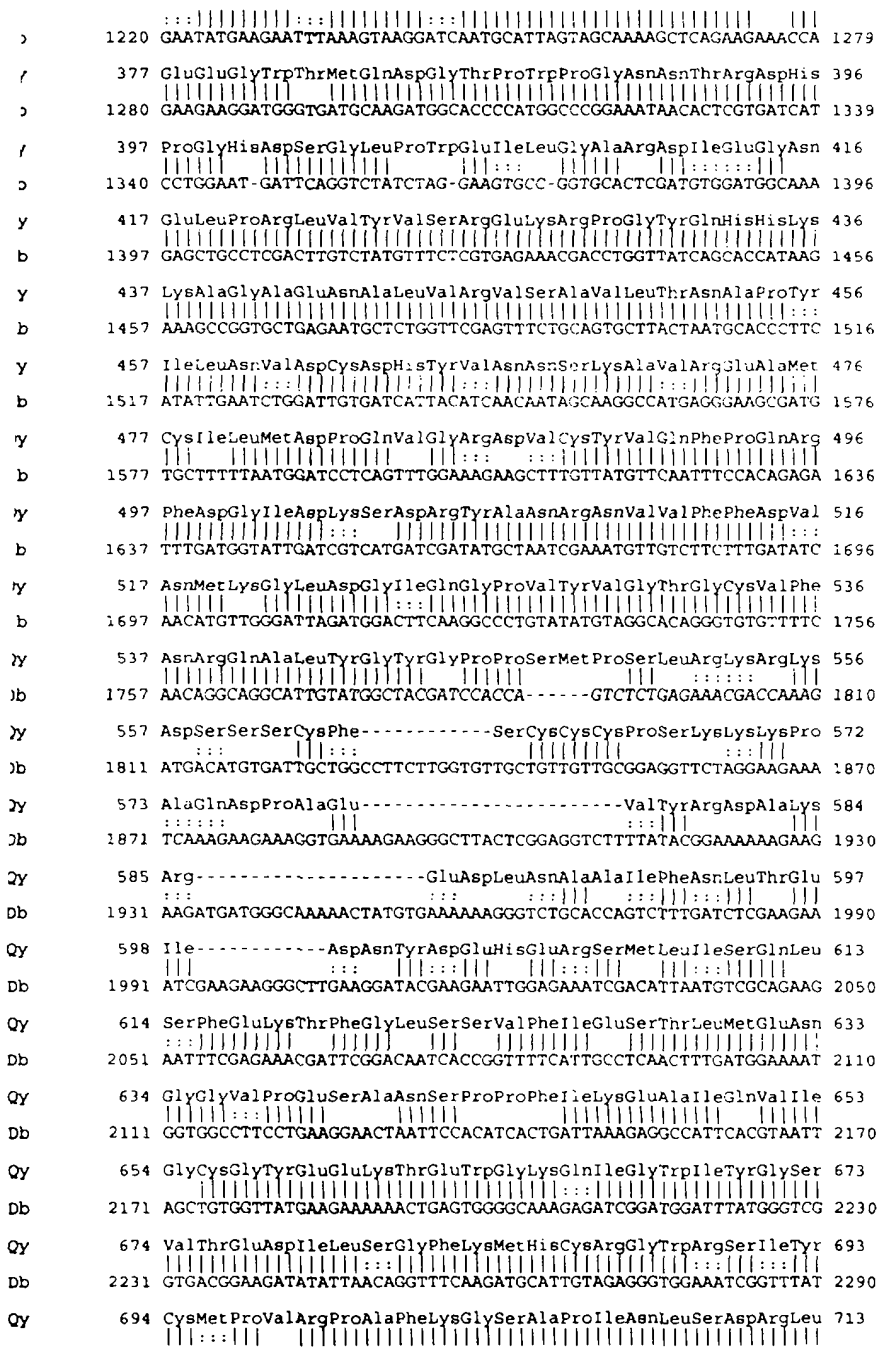
Qy        317 PheCysLysLysPheSerIleGluProArgAlaProGluPheTyrPheSerGlnLysIle 336

Db       1100 TTTTGTAAGAAGCATAATGTTGAGCCCAGGGCGCCGGAGTTTTATTTCAATGAGAAGATT 1159

Qy        337 AspTyrLeuLysAspLysValGlnProSerPheValLysGluArgArgAlaMetLysArg 356

Db       1160 GATTATTTGAAGGACAAGGTCCATCCTAGCTTTGTTAAAGAACGGAGAGCCATGAAAAGG 1219

Qy        357 AspTyrGluGluTyrLysValArgValAsnAlaLeuValAlaLysAlaGlnLysThrPro 376
```



```
Db       2291 TGTGTACCGAAAAGAACCGGCATTCAAAGGGTCCGCTCCAATCAATCTCTCGGATCGGTTG 2350

Qy        714 HisGlnValLeuArgTrpAlaLeuGlySerValGluIlePhePheSerArgHisCysPro 733

Db       2351 CACCAAGTTTTGAGATGGGCACTTGGTTCTGTAGAAATTTTCCTTAGTCGTCACTGTCCA 2410

Qy        734 LeuTrpTyrGlyPheGlyGlyGlyArgLeuLysTrpLeuGlnArgLeuAlaTyrIleAsn 753

Db       2411 CTTTGGTATGGTTA--GGTGGAAAACTGAAATGGCTCGAGAGGCTTGCTTATATCAAC 2467

Qy        754 ThrIleValTyrProPheThrSerLeuProLeuIleAlaTyrCysThrIleProAlaVal 773

Db       2468 ACCATTGTTTACCTTTCAACTCGATCCCTTTACTCGCCTATTGTACTATTCCAGCTGTT 2527

Qy        774 CysLeuLeuThrGlyLysPheIleIleProThrLeuSerAsnLeuAlaSerMetLeuPhe 793

Db       2528 TGTCTTCTCACCGGAAATTCATCATTCCAACTCTAAGCAACCTTACAAGTCTGTGGTTC 2587

Qy        794 LeuGlyLeuPheIleCysIleIleValThrAlaValLeuGluLeuArgTrpSerGlyVal 813

Db       2588 TTGGCACTCTTCTCATTATTTAAATCTTCTTTGAACTTCGATGGAGCGGGGTT 2647

Qy        814 SerIleGluAspLeuTrpArgAsnGluGlnPheTrpValIleGlyGlyValSerAlaHis 833

Db       2648 AGCATCCAAGACTGTGGCGCAATGAACAATTCTGGGTGATCGGAGGTGTCTCCGCCCAT 2707

Qy        834 LeuPheAlaValPheGlnGlyPheLeuLysMetLeuAlaGlyIleAspThrAsnPheThr 853

Db       2708 CTTTTTGCTGTCTTCAGGGCCTCCTCAAAGTCTAGCTGGAGTAGACACCAACTTCACC 2767

Qy        854 ValThrAlaLysAlaAlaGluAspAlaGluPheGlyGluLeuTyrMetValLysTrpThr 873

Db       2768 GTAACAGCAAAAGCAGCAGACGATACAGAATCTGGTGAACTTTATCTCTTCAAATGGACA 2827

Qy        874 ThrLeuLeuIleProProThrThrLeuLeuIleIleAsnMetSerGly---CysAlaGly 892

Db       2828 ACTCTCTTAATCCGTCCCACAACTCTGATAATACTGAACATGGTCGGAGTCGTGGCCGGA 2887

Qy        893 PheSerAspAlaLeuAsnLysGlyTyrGluAlaTrpGlyProLeuPheGlyLysValPhe 912

Db       2888 GTTTCAGACGCAATTAACAACGGCTATGGTTCATGGGGTCCATTGTTCGGCAAACTGTTC 2947

Qy        913 PheAlaPheTrpValIleLeuHisLeuTyrProPheLeuLysGlyLeuMetGlyArgGln 932

Db       2948 TTCGCATTCTGGGTCATTCTTCATCTTTACCCATTCCTCAAAGGTTTGATGGGGAGACAA 3007

Qy        933 AsnLeuThrProThrIleValValLeuTrpSerValLeuLeuAlaSerValPheSerLeu 952

Db       3008 AACAGGACGCCCACCATTGTTGTGCTTTGGTCCATACTTTGGCATCGATTTTCTCACTG 3067

Qy        953 ValTrpValLysIleAsnProPheValAsnLysValAspAsnThrLeuValAlaGluThr 972

Db       3068 GTTTGGGTACGGATTGATCCCTTCTTGCCCAAACAAACAGGTCCAGTTCTTAAACAA--- 3124

Qy        973 CysIleSerIleAsp-Cys 978

Db       3125 ---TGTGTCGTGGATTGC 3139

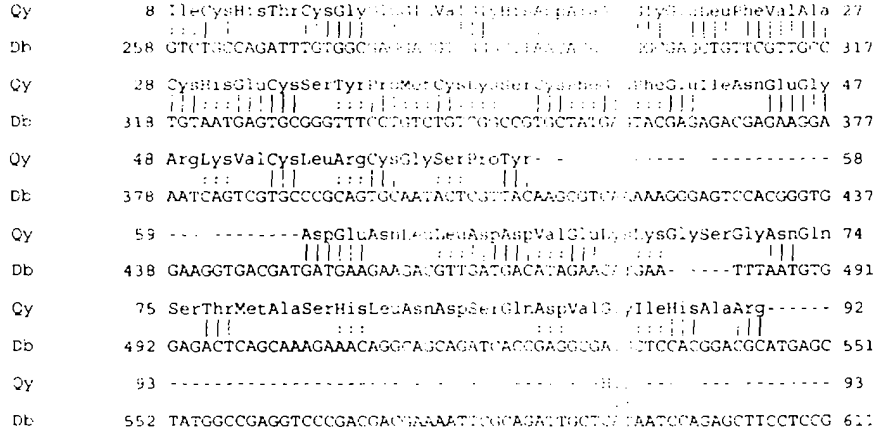
RESULT 6
AAA67114
ID   AAA67114 standard; DNA; 3951 BP.
XX
AC   AAA67114;
XX
DT   31-OCT-2000  (first entry)
XX
DE   Pinus radiata cellulose synthase nucleotide sequence SEQ ID NO:69.
XX
KW   Eucalyptus grandis; pinus radiata; Monterey pine; modification;
KW   plant cell wall; polysaccharide; polysaccharide biosynthetic pathway;
KW   transgenic plant; ds.
XX
OS   Pinus radiata.
```

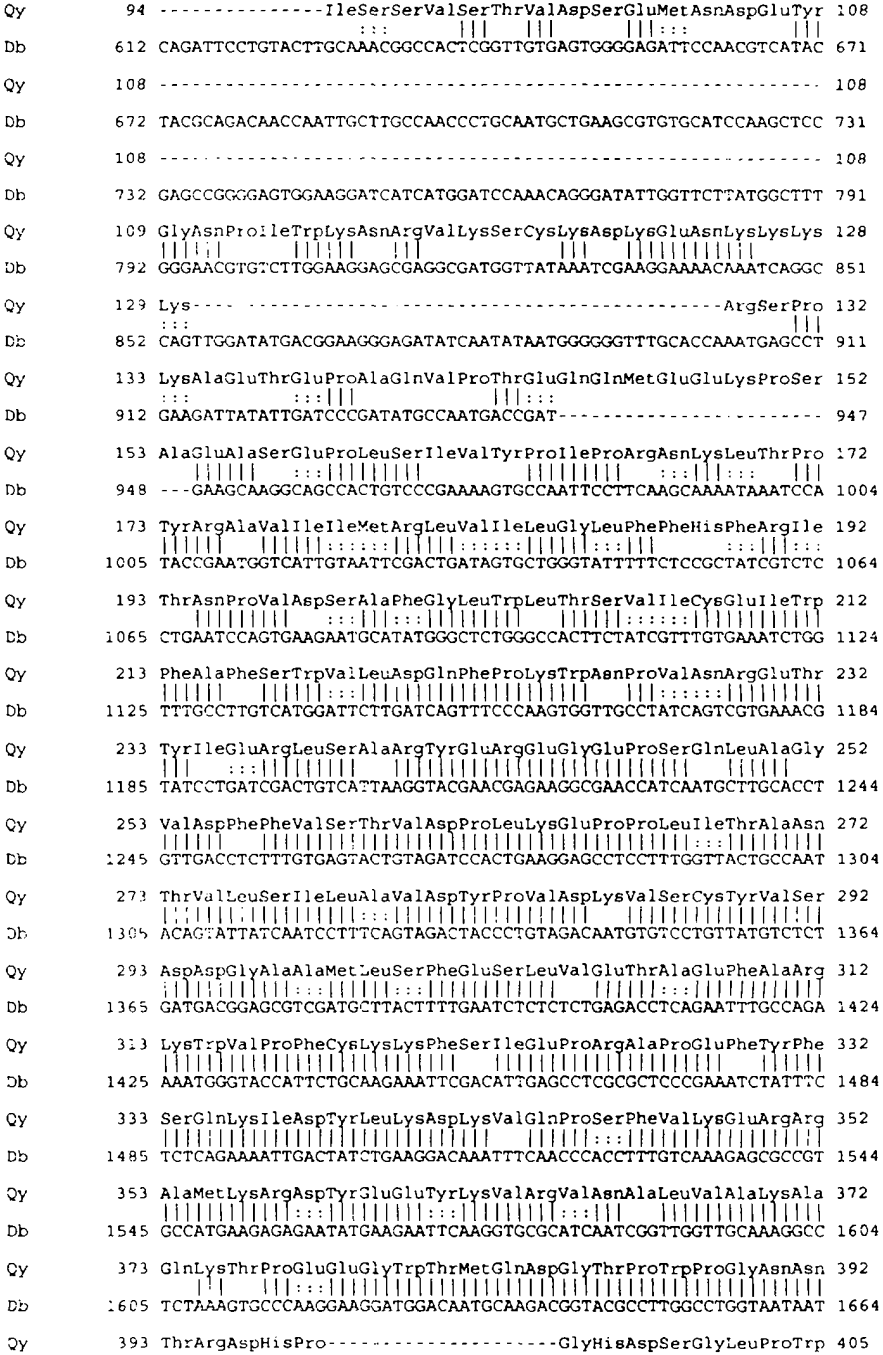
XX
PN WO200022092-A2.
XX
PD 20-APR-2000.
XX
PF 08-OCT-1999; 99WO-NZ00169.
XX
PR 13-OCT-1998; 98US-0170862.
PR 11-AUG-1999; 99US-0148426.
XX
PA (GENE-) GENESIS RES & DEV CORP LTD.
PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI Bloksberg LN;
XX
DR WPI; 2000-339328/29.
DR P-PSDB; AAB16307.
XX
PT New genes encoding proteins involved in a plant polysaccharide
PT biosynthetic pathway, useful for modulating or altering the
PT polysaccharide content, composition or structure of the plant -
XX
PS Claim 1; Page 71-72; 301pp; English.
XX
CC The present invention describes isolated polynucleotides (PN) comprising
CC a sequence selected from one of 835 nucleotide sequences given in
CC AAA67073 to AAA67907, their (reverse) complements, sequences producing
CC an Expectation (E) value of 0.01 or less compared to the 835 sequences,
CC sequences at least 50% identical to them, 200, 100, 40 or 20-mers of the
CC 835 sequences or sequences that are degenerately equivalent or allelic
CC to the 835 sequences. The polynucleotides are used to modify the
CC activity of a polypeptide involved in a polysaccharide biosynthetic
CC pathway in the plant. They are especially used to modulate or alter the
CC polysaccharide content, composition or structure of the plant. AAB16268
CC to AAB16340 are proteins encoded by some of the polynucleotide sequence
CC given in the present invention.
XX
SQ Sequence 3851 BP; 1040 A; 796 C; 944 G; 1071 T; 0 other;

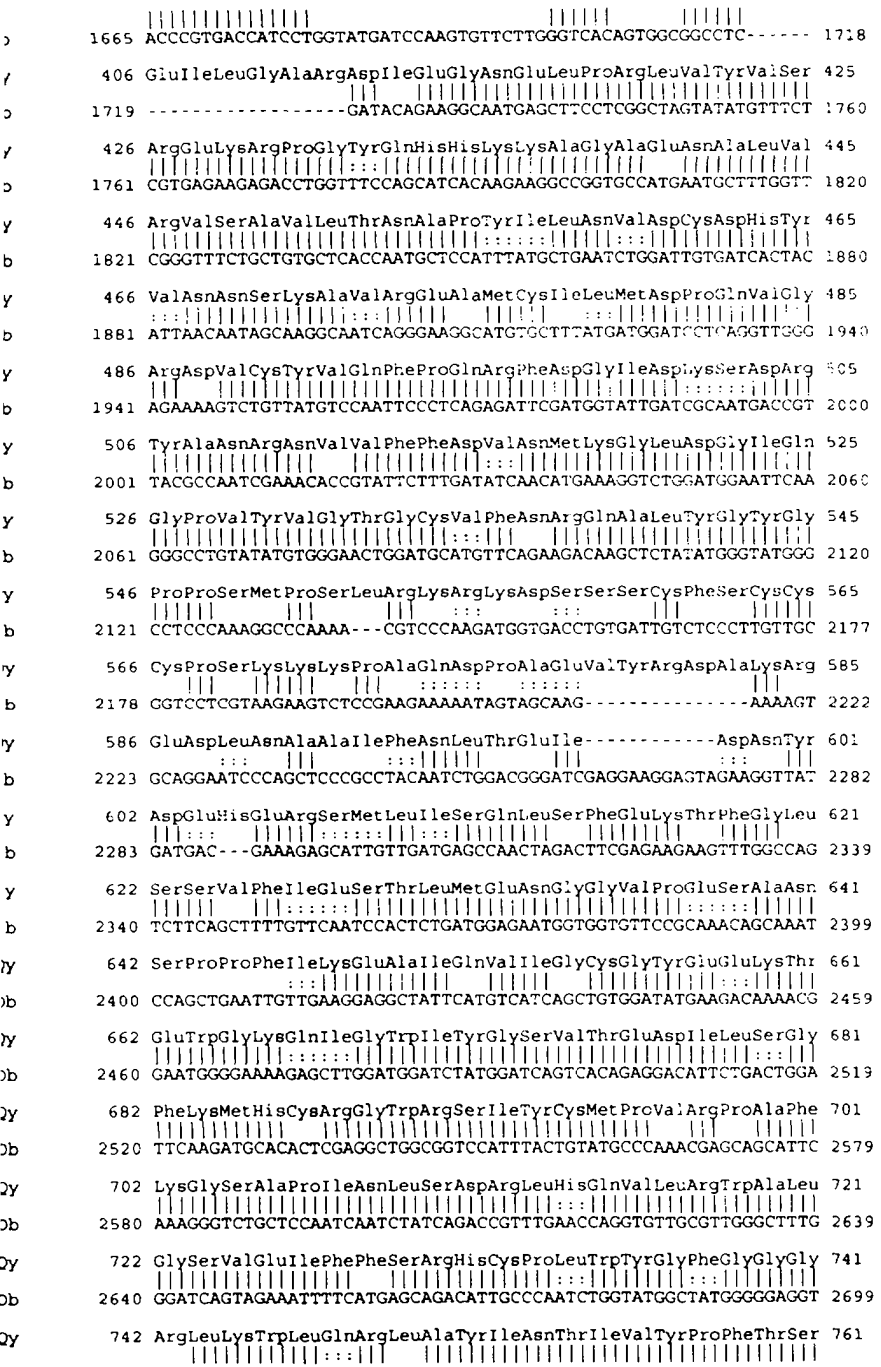
Alignment Scores:

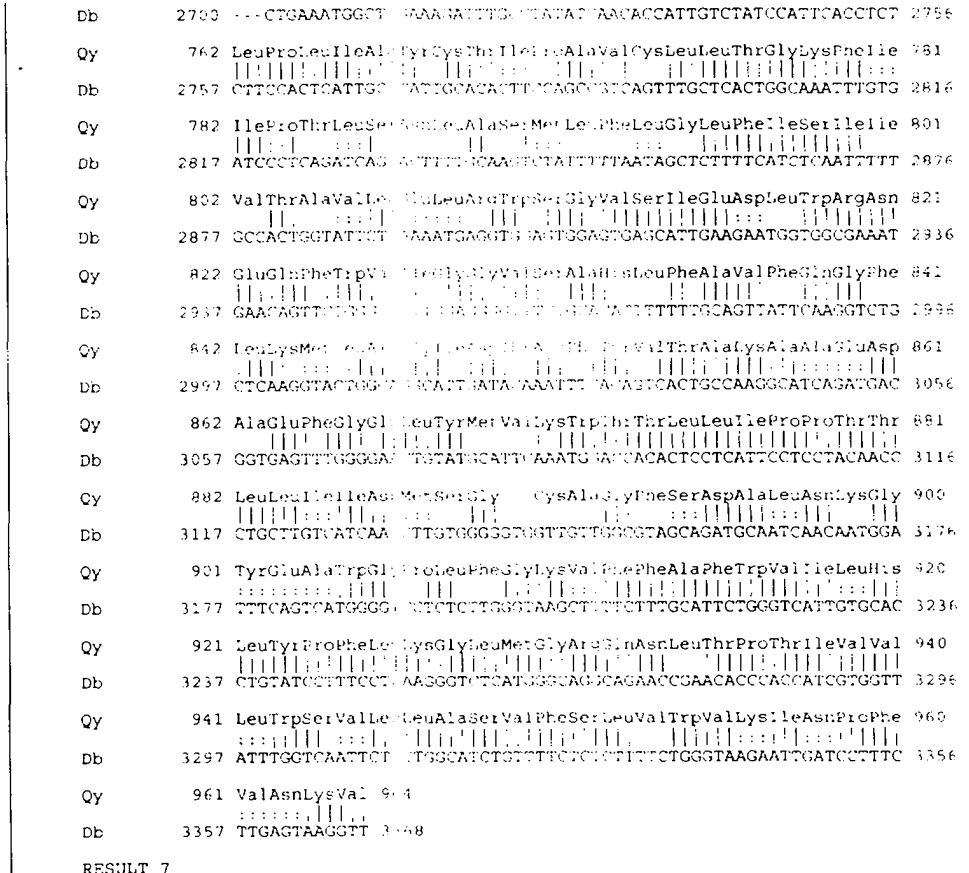
Pred. No.:	0	Length:	3851
Score:	3474.00	Matches:	1195
Percent Similarity:	72.46%	Conservative:	116
Best Local Similarity:	61.56%	Mismatches:	559
Query Match:	66.30%	Indels:	154
DB:	21	Gaps:	15

US-09-980-043-2 (1-978) x AAA67114 (1-3851)









RESULT 7
AAS16458
ID AAS16458 standard; cDNA. 3299 BP.
XX
AC AAS16458;
XX
DT 14-FEB-2002 (first entry)
XX
DE Corn cDNA encoding cellulose synthase Cqrael9/cesA-19.
XX
KW Corn; ss; cellulose synthase; Cdpgs45; cesA 3; Cqrael9; cesA-9;
KW stalk quality; improved stand; silage; perennarp; kernel hardening;
KW handling ability; transgenic plant.
XX
OS Zea mays.
XX
FH Key Location/Qualifiers
FT CDS 238..3299
FT /*tag= a
FT /product= "Cellulose synthase"
XX
PN WO200179516-A2.
XX
PD 25-OCT-2001.
XX
PF 12-APR-2001; 2001WO-US11951.
XX

PR 14-APR-2000; 2000US-0550483.
XX
PA (PION-) PIONEER HI-BRED INT INC.
XX
PI Dhugga KS, Helentjaris TG;
XX
DR WPI; 2002-041338/05.
DR P-PSDB; AAU10496.
XX
PT New cellulose synthase polypeptides and polynucleotides, useful in
PT improving stalk quality or silage, and in increasing concentration of
PT cellulose in the pericarp, hardening the kernel for improved handling
PT ability
XX
PS Claim 4; Page 80-85; 88pp; English.
XX
CC The invention relates to isolated nucleic acids encoding two cellulose
CC synthase proteins from corn, Cdpgs45 (cesA-3) and Cgrae19 (cesA-9).
CC Also disclosed are a recombinant expression cassette comprising the
CC polynucleotide (operably linked to a promoter) a host cell comprising
CC the recombinant expression cassette and a transgenic plant comprising
CC the recombinant expression cassette. The nucleic acid is useful in
CC the improvement of stalk quality for improved stand or silage, and in
CC the increased concentration of cellulose in the pericarp, hardening
CC the kernel, and thus improving its handling ability. The nucleic acids
CC may also be used as probes or amplification primers in the detection,
CC quantification or isolation of gene transcripts, as probes in detecting
CC deficiencies in the level of mRNA, for detecting gene mutations or
CC allelic variants, for monitoring up regulation of expression or changes
CC in enzyme activity in screening assays, for site directed mutagenesis,
CC and in sense or antisense suppression of one or more genes in a host
CC cell, tissue or plant. The polypeptides may be used in assays for
CC enzyme agonists or antagonists, and as immunogen or antigen to
CC obtain antibodies specifically immunoreactive with the protein.
CC The present sequence encodes a corn cellulose synthase of the
CC invention.
XX
SQ Sequence 3799 BP; 925 A; 931 C; 995 G; 944 T; 4 other;

Alignment Scores:

Pred. No.:	0	Length:	3799
Score:	3366.50	Matches:	648
Percent Similarity:	72.58%	Conservative:	143
Best Local Similarity:	60.22%	Mismatches:	193
Query Match:	64.25%	Indels:	114
DB:	24	Gaps:	17

US-09-980-043-2 (1 978) x AAS16458 (1 3799)

```
Qy          5 GlyAlaProIleCysHisThrCysGlyIleGlnValGlyLeuAspAlaAsnGlyGluLeu 24
              ||    :::||    ||  |  ||  |  ||    ||    |.:::::::
Db        280 GGAGGGCAGGTGTGCCAGATCTGCGGAATCGAGGTAGCTCTAGGCGAGGGAGACGTC 339

Qy         25 PheValAlaCysHisGluCysSerTyrProMetCysLysProCysPheGluPheGluIle 44
              ||  ||||||      |||  :::|:|:::| |:::  |':::::|:::|||
Db        340 TTCACCGCCTGCGACGTCTGCGGGTTCCCGGTGTGCCGCCCTGCTACGAGTACGAGCGC 399

Qy         45 AsnGluGlyArgLysValCysLeuArgCys --------  ----------GlySer 56
                 :::|||   :::   |||   :::|'|                    ||||||
Db        400 AAGGACGGCACACAAGCGTGCCCCCAGTGCAAAAAATAAGTAAAGCGCCACAAGGGGAGT 459

Qy         57 Pro---------------- -  --- -  -  -----  ---  ---------- 57
              |||
Db        460 CCAGCGATCCGAGGGGAGGAAGGAGACGATACTGATGCCGAGATGCTAGCGACTTCAAC 519

Qy         58 TyrAspGluAsnLeuLeuAspAspValGluLysLysGlySerGlyAsnGlnSerThrMet 77
              ||'       :::     |||:||   :::::|||   ::              :::
Db        520 TACCCTGCATCTGGCAATGACGACCAGAAGCAGAAGATTGCGACAGGATGCGCAGCTGG 579

Qy         78 AlaSerHisLeuAsnAspSerGlnAspValGly  - -   ---  ------------ 88
                 :::         |||   ||:|:|||'
Db        580 CGCATGAATGCTGGGGGCAGCGGGGATGTTGGCCGCCCCAA'ATGACAGTGGTGAGATC 639

Qy         89 ---------------------------IleHisAlaArgHisIleSerSerVal------ 97
                                        |||         :::|||   ||||||
Db        640 GGGCTTACCAAGTACGACAGTGGTGAGATCCCTCGGGGATACATCCCGTCAGTCACTAAC 699

Qy         97 ------------------------------------------------------------ 97

Db        700 AGCCAGATTTCGGGAGAAATCCCTGGTGCTTCCCCTGACCATCATATGATGTCTCCTACT 759

Qy         98 ---------------------------------------------SerThrValAspSer 102
                                                          |||
Db        760 GGGAACATTGGCAGGCGCGCCCCATTTCCCTATATGAATCATTCATCAAATCCGTCGAGG 819

Qy        103 GluMetAsnAspGluTyrGlyAsnProIleTrpLysAsnArgValLysSerCysLysAsp 122
              |||   :::          ||||||      ||||||  ||||||          |||
Db        820 GAATTCTCTGGTAGCGTTGGGAATGTTGCCTGGAAAGAGAGGGTTGATGGCTGGAAAATG 879

Qy        123 LysGluAsnLys------------LysLysLysArgSerProLysAlaGluThrGluPro 138
              |||:::::|||                    :::      |||   :::|||
Db        880 AAGCAGGACAAGGGAACAATTCCCATGACGAATGGCACAAGCATTGCTCCCTCTGAGGGC 939

Qy        139 AlaGlnVal--------------ProThrGluGlnGlnMetGluGluLys---ProSer 152
                  |||              |||:::     |||||||:::         :::
Db        940 CGGGGTGTTGGTGATATTGATGCATCAACTGATTACAACATGGAAGATGCCTTATTAAAC 999

Qy        153 AlaGluAlaSerGluProLeuSerIleValTyrProIleProArgAsnLysLeuThrPro 172
                 |||      :::|||||||||         |||:::|||   :::::::::   |||
Db       1000 GATGAAACTCGCCAGCCTCTATCTAGGAAAGTTCCACTTCCTTCCTCCAGGATAAATCCA 1059

Qy        173 TyrArgAlaValIleIleMetArgLeuValIleLeuGlyLeuPhePheHisPheArgIle 192
              |||||   ||||||:::::|||||:::::|||   :::|||   |||:::|||||
Db       1060 TACAGGATGGTCATTGTGCTACGATTGATTGTTCTAAGCATCTTCTTGCACTACCGGATC 1119

Qy        193 ThrAsnProValAspSerAlaPheGlyLeuTrpLeuThrSerValIleCysGluIleTrp 212
              ||||||||||||   :::|||:::   |||||||||   ||||||||||||||||||||
Db       1120 ACAAATCCTGTGCGTAATGCATACCCACTGTGGCTTCTATCTGTTATATGTGAGATCTGG 1179

Qy        213 PheAlaPheSerTrpValLeuAspGlnPheProLysTrpAsnProValAsnArgGluThr 232
              ||||||   |||||:::|||||||||||||||||||||||   |||:::||||||||||
Db       1180 TTTGCTCTTTCCTGGATATTGGATCAGTTTCCAAAGTGGTTTCCAATCAACCGCGAGACT 1239

Qy        233 TyrIleGluArgLeuSerAlaArgTyrGluArgGluGlyGluProSerGlnLeuAlaGly 252
              |||:::::|||||:::   |||||||:::|||||||||||||||||||||||||||||
Db       1240 TACCTTGATAGACTCGCATTAAGGTATGACCGGGAAGGTGAGCCATCTCAGTTGGCTGCT 1299

Qy        253 ValAspPhePheValSerThrValAspProLeuLysGluProProLeuIleThrAlaAsn 272
              ||||||   ||||||||||||||||||||:::|||||||||||||||:::|||||||||
Db       1300 GTTGACATTTTTGTCAGTACTGTCGACCCAATGAAGGAGCCTCCTCTTGTCACTGCCAAT 1359

Qy        273 ThrValLeuSerIleLeuAlaValAspTyrProValAspLysValSerCysTyrValSer 292
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1360 ACCGTGCTATCCATTCTCGCTGTGGACTATCCTGTGGATAAGGTCTCTTGCTATGTATCT 1419

Qy        293 AspAspGlyAlaAlaMetLeuSerPheGluSerLeuValGluThrAlaGluPheAlaArg 312
              |||||||||||||||||||||:::|||:::::|||   ||||||:::|||||||||||
Db       1420 GATGATGGAGCTGCTATGCTGACATTTGATGCACTAGCTGAGACTTCAGAGTTTGCTAGA 1479

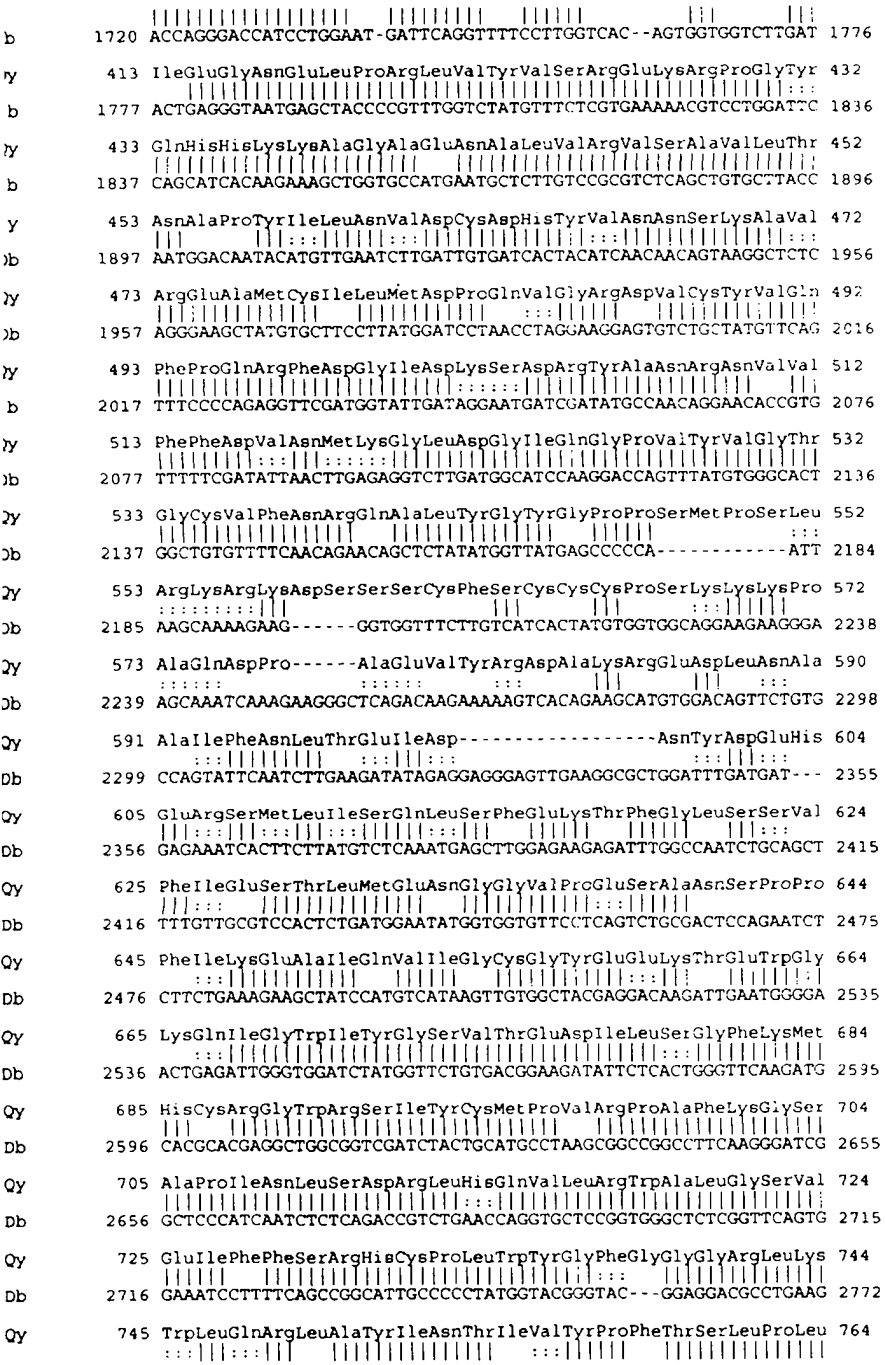
Qy        313 LysTrpValProPheCysLysLysPheSerIleGluProArgAlaProGluPheTyrPhe 332
              |||||||||||||||   ||||||:::::|||||||||||||||||||||:::||||||
Db       1480 AAATGGGTGCCATTTGTTAAGAAGTACAACATTGAACCTAGAGCTCCTGAATGGTACTTC 1539

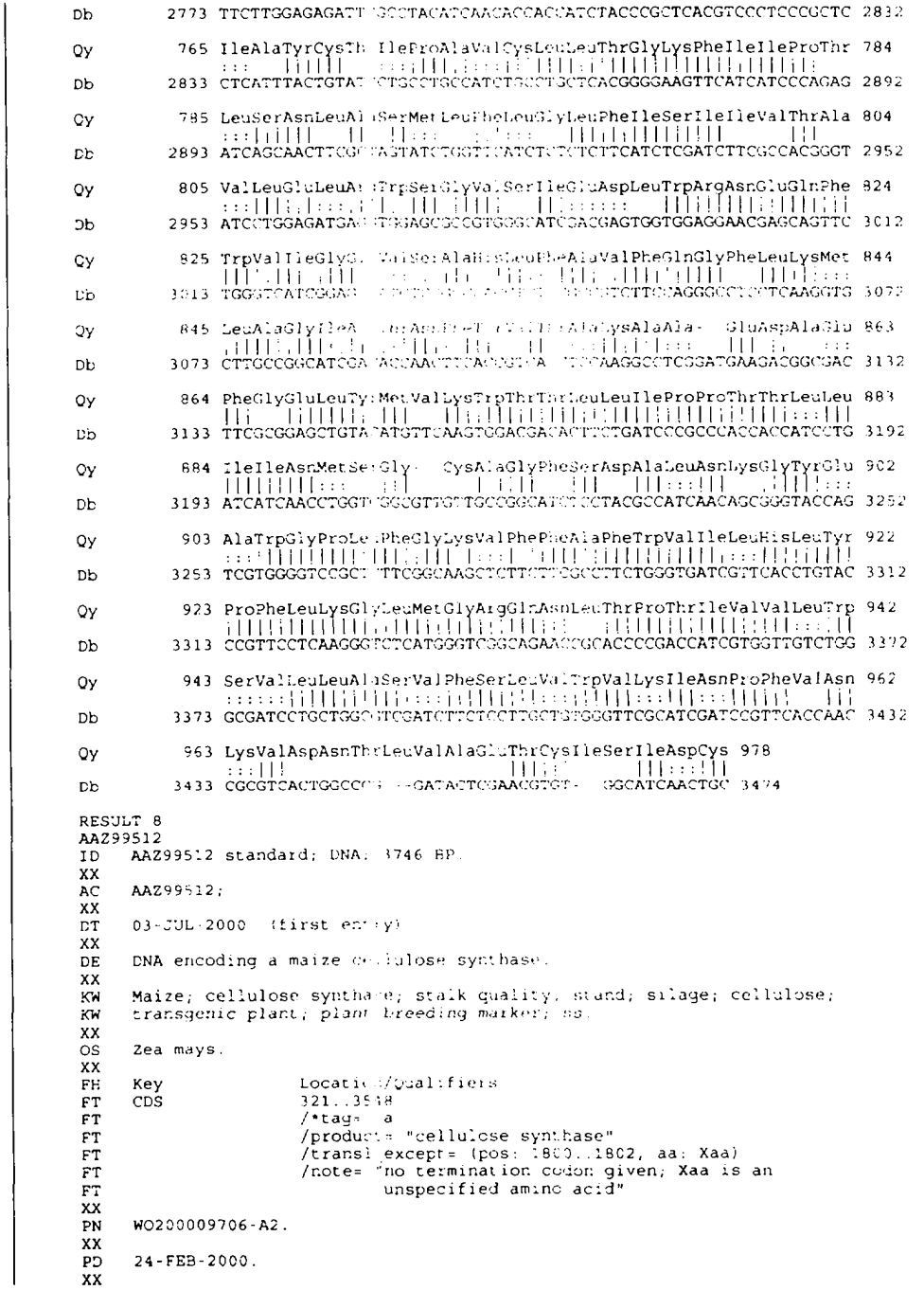
Qy        333 SerGlnLysIleAspTyrLeuLysAspLysValGlnProSerPheValLysGluArgArg 352
              |||||||||||||||||||||||||||||||||   |||||||||||||||:::||||||
Db       1540 TCCCAGAAAATTGATTACTTGAAGGACAAAGTGCACCCTTCATTTGTTAAAGACCGCCGG 1599

Qy        353 AlaMetLysArgAspTyrGluGluTyrLysValArgValAsnAlaLeuValAlaLysAla 372
              ||||||||||||:::|||||||||:::|||:::|||||||||   ||||||||||||||
Db       1600 GCCATGAAGAGAGAATATGAAGAATTCAAAATTAGGGTAAATGGCCTTGTTGCTAAGGCA 1659

Qy        373 GlnLysThrProGluGluGlyTrpThrMetGlnAspGlyThrProTrpProGlyAsnAsn 392
              ||||||   ||||||||||||||   |||||||||||||||||||||||||||||||||
Db       1660 CAAAAAGTCCCTGAGGAAGGATGGATCATGCAAGATGGCACACCATGGCCAGGAAACAAT 1719

Qy        393 ThrArgAspHisProGlyHisAspSerGlyLeuProTrpGluIleLeuGlyAlaArgAsp 412
```





RESULT 8
AAZ99512
ID AAZ99512 standard; DNA; 3746 BP.
XX
AC AAZ99512;
XX
DT 03-JUL-2000 (first entry)
XX
DE DNA encoding a maize cellulose synthase.
XX
KW Maize; cellulose synthase; stalk quality, stand; silage; cellulose;
KW transgenic plant; plant breeding marker; ss.
XX
OS Zea mays.
XX
FH Key Location/Qualifiers
FT CDS 321..3518
FT /tag= a
FT /product= "cellulose synthase"
FT /transl_except= (pos:1800..1802, aa:Xaa)
FT /note= "no termination codon given; Xaa is an
FT unspecified amino acid"
XX
PN WO200009706-A2.
XX
PD 24-FEB-2000.
XX

```
PF   16 AUG-1999;   99WO-US18760.
XX
PR   17-AUG-1998;   98US-0096822.
XX
PA   (PION-) PIONEER HI-BRED INT INC.
XX
PI   Dhugga KS,  Helentjaris TG,  Bowen BA,  Wang X;
XX
DR   WPI; 2000-224343/19.
DR   P-PSDB; AAY84114.
XX
PT   New genes which encode maize cellulose synthase polypeptides in plants
PT   useful for modulating the expression of cellulose synthase in plants
PT   and to produce transgenic plants expressing the novel protein
XX
PS   Claim 1; Page 137-141; 119pp; English
XX
CC   The present sequence encodes a maize cellulose synthase polypeptide.
CC   The cellulose synthase can be used for the improvement of stalk quality
CC   for improved stand or silage. It also provides an increased concentration
CC   of cellulose in the pericarp, hardening the kernel and improving its
CC   handling ability. The sequences are used to produce transgenic plants
CC   and seeds expressing the cellulose synthase. The polynucleotide is
CC   used for modulating, preferably increasing, the level of the synthase
CC   in a plant cell. The plants are preferably monocots. The polynucleotide
CC   is also used as a probe or primer in the detection quantitation or
CC   isolation of gene transcripts. The probes are useful in detecting
CC   deficiencies in the level of mRNA in screenings for desired transgenic
CC   plant, for detecting mutations in the gene, for monitoring upregulation
CC   of expression or changes in enzyme activity in screening assays of
CC   compounds, for detection of any number of allelic variants of the gene,
CC   or for use as molecular markers in plant breeding programs. The
CC   isolated nucleic acids of the present invention can also be used for
CC   recombinant expression of their encoded polypeptides or for use as
CC   immunogens in the preparation and/or screening of antibodies. The
CC   proteins can be employed in assays for enzyme agonists or antagonists
CC   of enzyme function or for use of immunogens or antigens to obtain
CC   antibodies specifically immunoreactive with a protein.
XX
SQ   Sequence 3746 BP; 861 A; 976 C; 1006 G; 902 T; 1 other;

Alignment Scores:
Pred. No.:              0              Length:          3746
Score:                  3365.50        Matches:         [illegible]
Percent Similarity:     73.00%         Conservative:    [illegible]
Best Local Similarity:  60.34%         Mismatches:      [illegible]
Query Match:            64.23%         Indels:          [illegible]
DB:                     21             Gaps:            [illegible]

US-09-980-043-2 (1-978) x AAZ99513 (1-3746)

Qy          5 GlyAlaProIleCysHisThrCysGlyGluGlnValGly[illegible]AspAlaAsnGlyGluLeu 24
              |||      :::|||      ||'|||:::  |||.||      |    .|||:::::
Db        363 GGCGGACAGGTGTGCCAGATCTGCGGCGACGGCGTGGGCAC[illegible]ACGGCGGAGGGGGACGTC 422

Qy         25 PheValAlaCysHisGluCysSerTyrProMetCysLysS[illegible]CysPheGluPheGluIle 44
              |||   ||||||     ||'   :::'|||:: ||:::   ||'|:::|||:::|||
Db        423 TTCGCCGCCTGCGACGTCTGCGGGTTTCCGGTGTGCCGC[illegible]TGCTACGAGTACGAGCGC 482

Qy         45 AsnGluGlyArgLysValCysLeuArgCys[illegible]----------GlySer 56
                 :::|||   :::   |||   :::|||                    |||||||
Db        483 AAGGACGGCACGCAGGCGTGCCCC[illegible]CAGTGCAAGACCAAGTA[illegible]AAGCGCCACAAGGGGAGC 542

Qy         57 Pro---------------[illegible]TyrAspGluAsnLeuLeu 63
              |||                                  |||   |||   |||
Db        543 CCGGCGATCCGTGGGGAGGAAGGAGACGACACTGATGCCGA[illegible]AGCGACTTCAATTACCTT 602

Qy         64 ------------AspAspValGluLysLysGlySerGlyAs[illegible]GlnSerThrMetAlaSer 79
                         :::|||   ::::::|||   :::          :::
Db        603 GCATCTGGCAATGAGGACCAGAAGCAGAAGATTGCCGACAGAATGCGCAGCTGGCGCATG 662

Qy         80 HisLeuAsnAspSerGlnAspValGly[illegible]----------------- 88
              ::::::      |||   ||||'||||

Db        663 AACGTTGGGGGCAGCGGGGATGTTGGTCGCCCCAAGTATGACAGTGGCGAGATCGGGCTT 722

Qy         89 --------------------IleHisAlaArgHisIleSerSerValSer--------- 98
                                  |||      :::|||   ||||||:::
Db        723 ACCAAGTATGACAGTGGCGAGATTCCTCGGGGATACATCCCATCAGTCACTAACAGCCAG 782

Qy         98 ------------------------------------------------------------ 98

Db        783 ATCTCAGGAGAAATCCCTGGTGCTTCCCCTGACCATCATATGATGTCCCCAACTGGGAAC 842

Qy         99 -------------------------------------------ThrValAspSerGluMet 104
                                                         |||
Db        843 ATTGGCAAGCGTGCTCCATTTCCCTATGTGAACCATTCGCCAAATCCGTCAAGGGAGTTC 902

Qy        105 AsnAspGluTyrGlyAsnProIleTrpLysAsnArgValLysSerCysLysAspLysGlu 124
              :::         ||||||      ||||||   ||||||          |||   |||:::
Db        903 TCTGGTAGCATTGGGAATGTTGCCTGGAAAGAGAGGGTTGATGGCTGGAAAATGAAGCAG 962

Qy        125 AsnLys------------LysLysLysArgSerProLysAlaGluThrGluProAlaGln 140
              :::|||                        :::      |||   :::|||
Db        963 GACAAGGGGACGATTCCCATGACGAATGGCACAAGCATTGCTCCCTCTGAGGGTCGGGGT 1022

Qy        141 Val---------------ProThrGluGlnGlnMetGluGluLys---ProSerAlaGlu 154
              |||                   |||:::      ||||||:::          :::   |||
Db       1023 GTTGGTGATATTGATGCATCAACTGATTACAACATGGAAGATGCCTTATTGAACGACGAA 1082

Qy        155 AlaSerGluProLeuSerIleValTyrProIleProArgAsnLysLeuThrProTyrArg 174
                  :::|||||||||        |||:::|||   :::::::::   ||||||||||
Db       1083 ACTCGACAGCCTCTATCTAGGAAAGTTCCACTTCCTTCCTCCAGGATAAATCCATACAGG 1142

Qy        175 AlaValIleIleMetArgLeuValIleLeuGlyLeuPhePheHisPheArgIleThrAsn 194
                |||||:::::|||||:::::::|||   :::|||   |||:::||||||||||||
Db       1143 ATGGTCATTGTGCTGCGATTGATTGTTCTAAGCATCTTCTTGCACTACCGTATCACAAAT 1202

Qy        195 ProValAspSerAlaPheGlyLeuTrpLeuThrSerValIleCysGluIleTrpPheAla 214
              ||||||   :::|||:::   |||||||||  ||||||||||||||||||||||||||||
Db       1203 CCTGTGCCGCAATGCATACCCATTATGGCTTCTATCTGTTATATGTGAGATCTGGTTTGCT 1262

Qy        215 PheSerTrpValLeuAspGlnPheProLysTrpAsnProValAsnArgGluThrTyrIle 234
                ||||||:::|||||||||||||||||||||   |||:::|||||||||||||||:::
Db       1263 CTTTCGTGGATATTGGATCAGTTCCCTAAGTGGTTTCCAATCAACCGGGAGACGTACCTT 1322

Qy        235 GluArgLeuSerAlaArgTyrGluArgGluGlyGluProSerGlnLeuAlaGlyValAsp 254
              :::|||||:::   ||||||:::|||||||||||||||||||||||||||   ||||||
Db       1323 GATAGGCTGGCATTAAGGTATGACCGGGAAGGTGAGCCATCTCAGTTGGCTGCTGTTGAC 1382

Qy        255 PhePheValSerThrValAspProLeuLysGluProProLeuIleThrAlaAsnThrVal 274
              |:|||||||||:|||||||:::|||:|||||||||||:::|||||||||||||||||
Db       1383 ATTTTCGTCAGTACAGTCGACCCAATGAAGGAGCCTCCTCTTGTCACTGCCAATACCGTG 1442

Qy        275 LeuSerIleLeuAlaValAspTyrProValAspLysValSerCysTyrValSerAspAsp 294
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1443 CTATCCATTCTTGCTGTGGATTACCCTGTGGATAAGGTCTCTTGCTATGTATCTGATGAT 1502

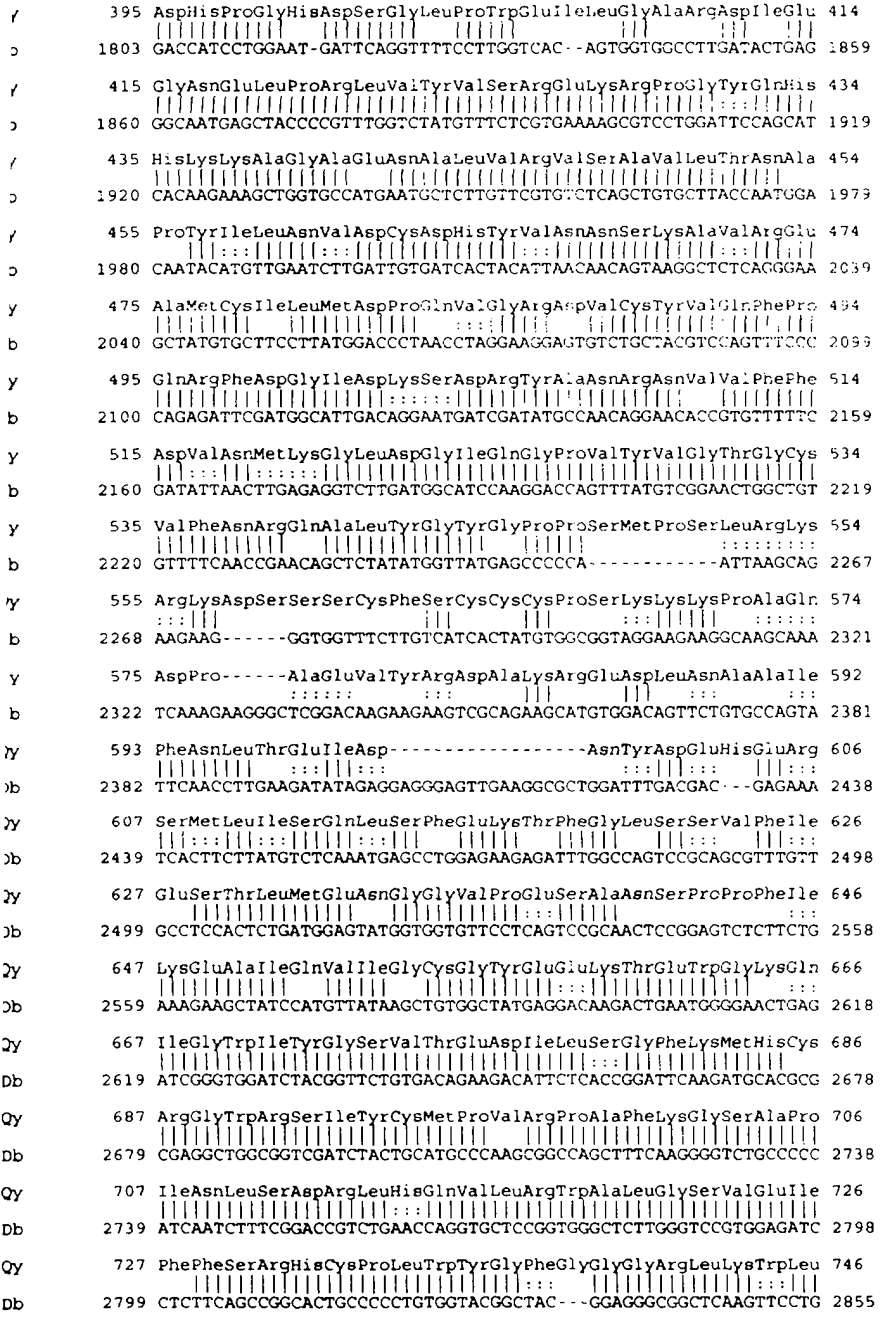
Qy        295 GlyAlaAlaMetLeuSerPheGluSerLeuValGluThrAlaGluPheAlaArgLysTrp 314
              ||||||||||||||:::|||:::::|||   ||||||:::|||||||||||||||||||
Db       1503 GGAGCTGCGATGCTGACATTTGATGCACTAGCTGAGACTTCAGAGTTTGCTAGAAAATGG 1562

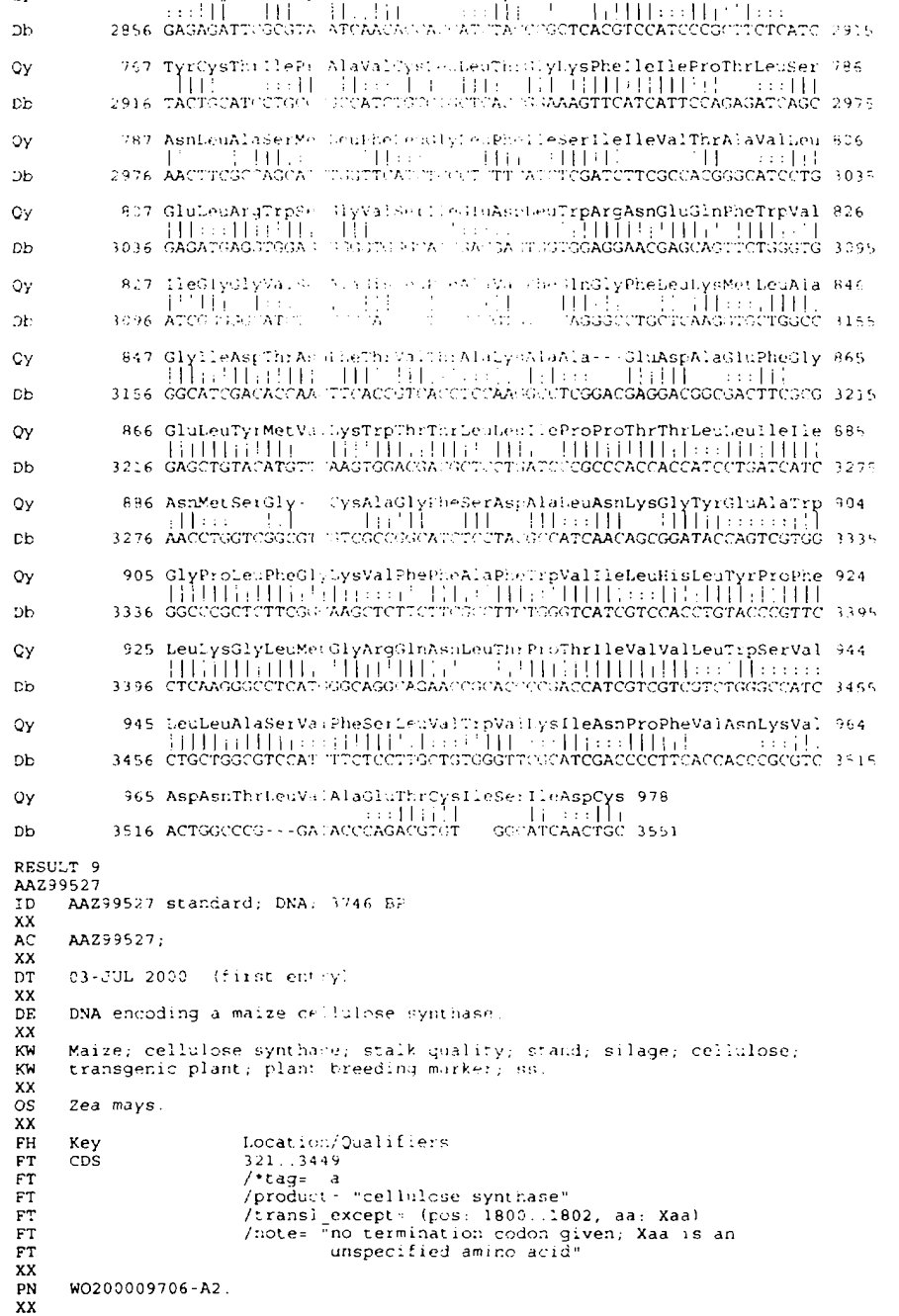
Qy        315 ValProPheCysLysLysPheSerIleGluProArgAlaProGluPheTyrPheSerGln 334
              |||||||||  |||||:::::|||||||||||||||||||||:::|||||||||||||
Db       1563 GTACCATTTGTTAAGAAGTACAACATTGAACCTAGAGCTCCTGAATGGTACTTCTCCCAG 1622

Qy        335 LysIleAspTyrLeuLysAspLysValGlnProSerPheValLysGluArgArgAlaMet 354
              ||||||||||||||||||||||||||   |||||||||||||:::|||||||||||||
Db       1623 AAAATTGATTACTTGAAGGACAAAGTGCACCCTTCATTTGTTAAAGACCGCCGGGCCATG 1682

Qy        355 LysArgAspTyrGluGluTyrLysValArgValAsnAlaLeuValAlaLysAlaGlnLys 374
              ||||||:::|||||||||.::|||||||||||||||   ||||||||||||||||||||
Db       1683 AAGAGAGAATATGAAGAATTCAAAGTTAGGGTAAATGGCCTTGTTGCTAAGGCACAGAAA 1742

Qy        375 ThrProGluGluGlyTrpThrMetGlnAspGlyThrProTrpProGlyAsnAsnThrArg 394
              |||||||||||||||   ||||||||||||||||||||||||||||||||||||||||
Db       1743 GTTCCTGAGGAAGGATGGATCATGCAAGATGGCACACCATGGCCAGGAAACAATACCMGG 1802
```





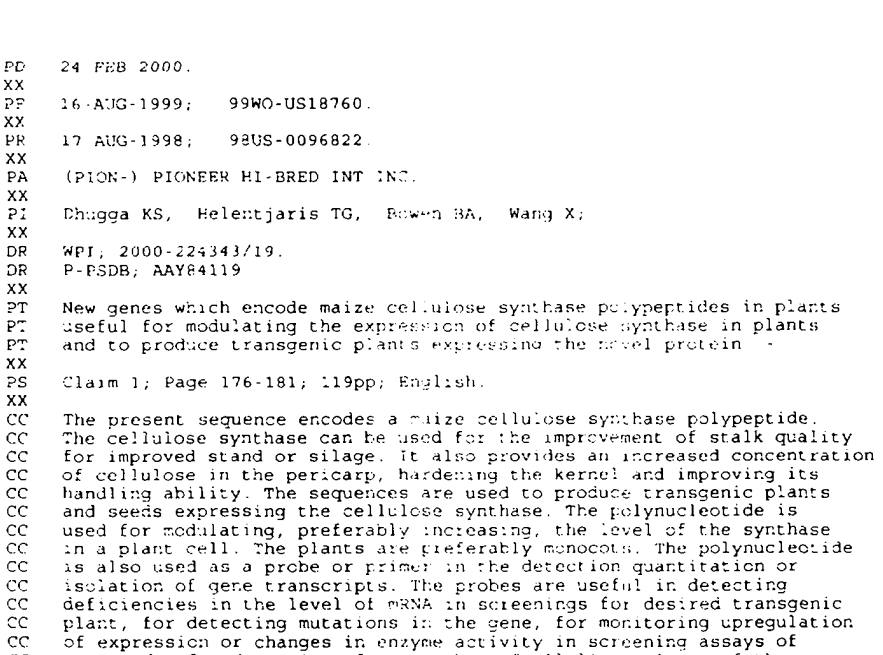
RESULT 9
AAZ99527
ID AAZ99527 standard; DNA; 3746 BP
XX
AC AAZ99527;
XX
DT 03-JUL 2000 (first entry)
XX
DE DNA encoding a maize cellulose synthase.
XX
KW Maize; cellulose synthase; stalk quality; stand; silage; cellulose;
KW transgenic plant; plant breeding marker; ss.
XX
OS Zea mays.
XX
FH Key Location/Qualifiers
FT CDS 321..3449
FT /*tag= a
FT /product= "cellulose synthase"
FT /transl_except= (pos: 1800..1802, aa: Xaa)
FT /note= "no termination codon given; Xaa is an
FT unspecified amino acid"
XX
PN WO200009706-A2.
XX

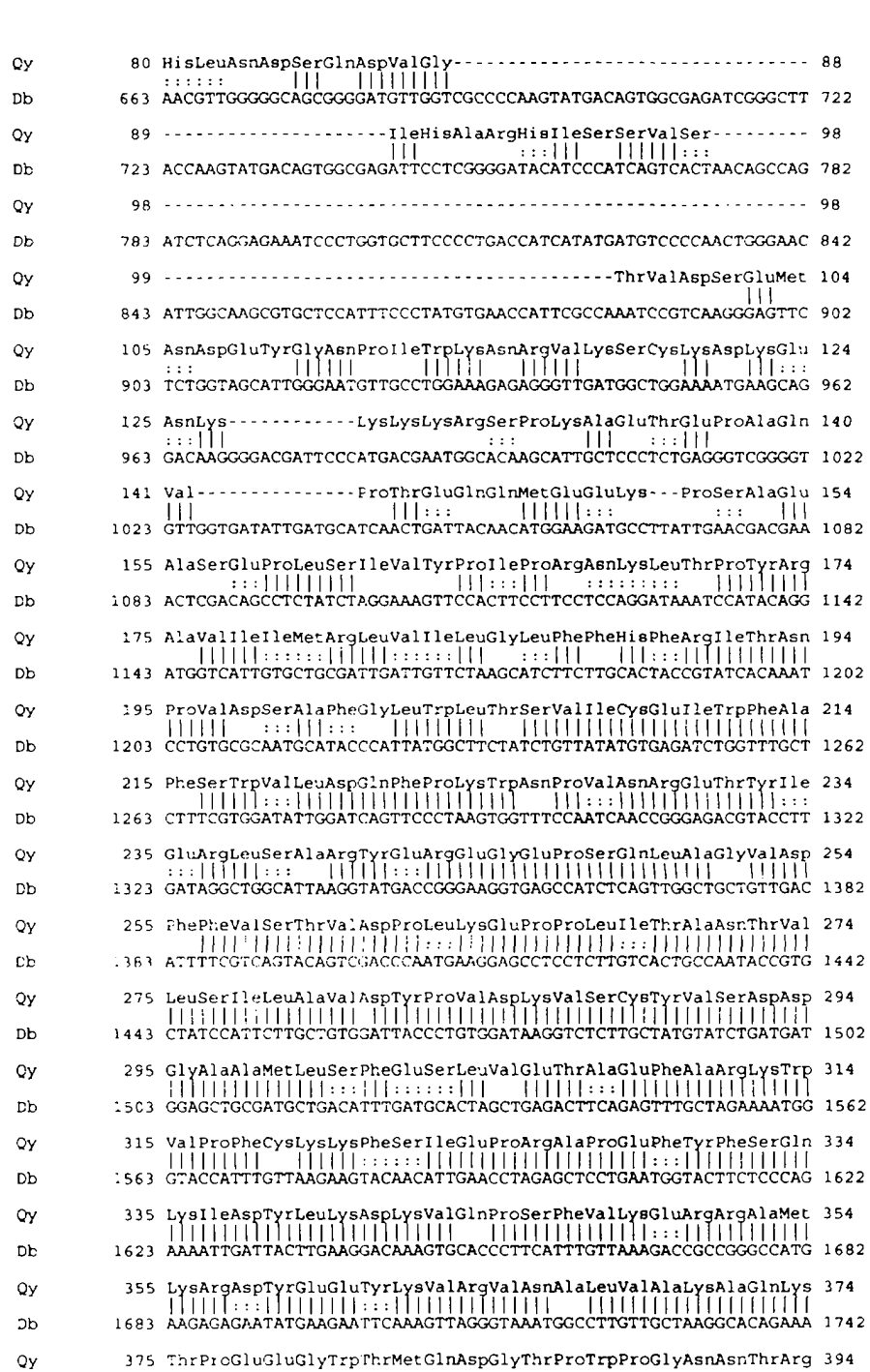
PD 24 FEB 2000.
XX
PF 16-AUG-1999; 99WO-US18760.
XX
PR 17-AUG-1998; 98US-0096822.
XX
PA (PION-) PIONEER HI-BRED INT INC.
XX
PI Dhugga KS, Helentjaris TG, Bowen BA, Wang X;
XX
DR WPI; 2000-224343/19.
DR P-PSDB; AAY84119
XX
PT New genes which encode maize cellulose synthase polypeptides in plants
PT useful for modulating the expression of cellulose synthase in plants
PT and to produce transgenic plants expressing the novel protein -
XX
PS Claim 1; Page 176-181; 119pp; English.
XX
CC The present sequence encodes a maize cellulose synthase polypeptide.
CC The cellulose synthase can be used for the improvement of stalk quality
CC for improved stand or silage. It also provides an increased concentration
CC of cellulose in the pericarp, hardening the kernel and improving its
CC handling ability. The sequences are used to produce transgenic plants
CC and seeds expressing the cellulose synthase. The polynucleotide is
CC used for modulating, preferably increasing, the level of the synthase
CC in a plant cell. The plants are preferably monocots. The polynucleotide
CC is also used as a probe or primer in the detection quantitation or
CC isolation of gene transcripts. The probes are useful in detecting
CC deficiencies in the level of mRNA in screenings for desired transgenic
CC plant, for detecting mutations in the gene, for monitoring upregulation
CC of expression or changes in enzyme activity in screening assays of
CC compounds, for detection of any number of allelic variants of the gene,
CC or for use as molecular markers in plant breeding programs. The
CC isolated nucleic acids of the present invention can also be used for
CC recombinant expression of their encoded polypeptides or for use as
CC immunogens in the preparation and/or screening of antibodies. The
CC proteins can be employed in assays for enzyme agonists or antagonists
CC of enzyme function or for use of immunogens or antigens to obtain
CC antibodies specifically immunoreactive with a protein.
XX
SQ Sequence 3746 BP; 861 A; 976 C; 1006 G; 902 T; 1 other;

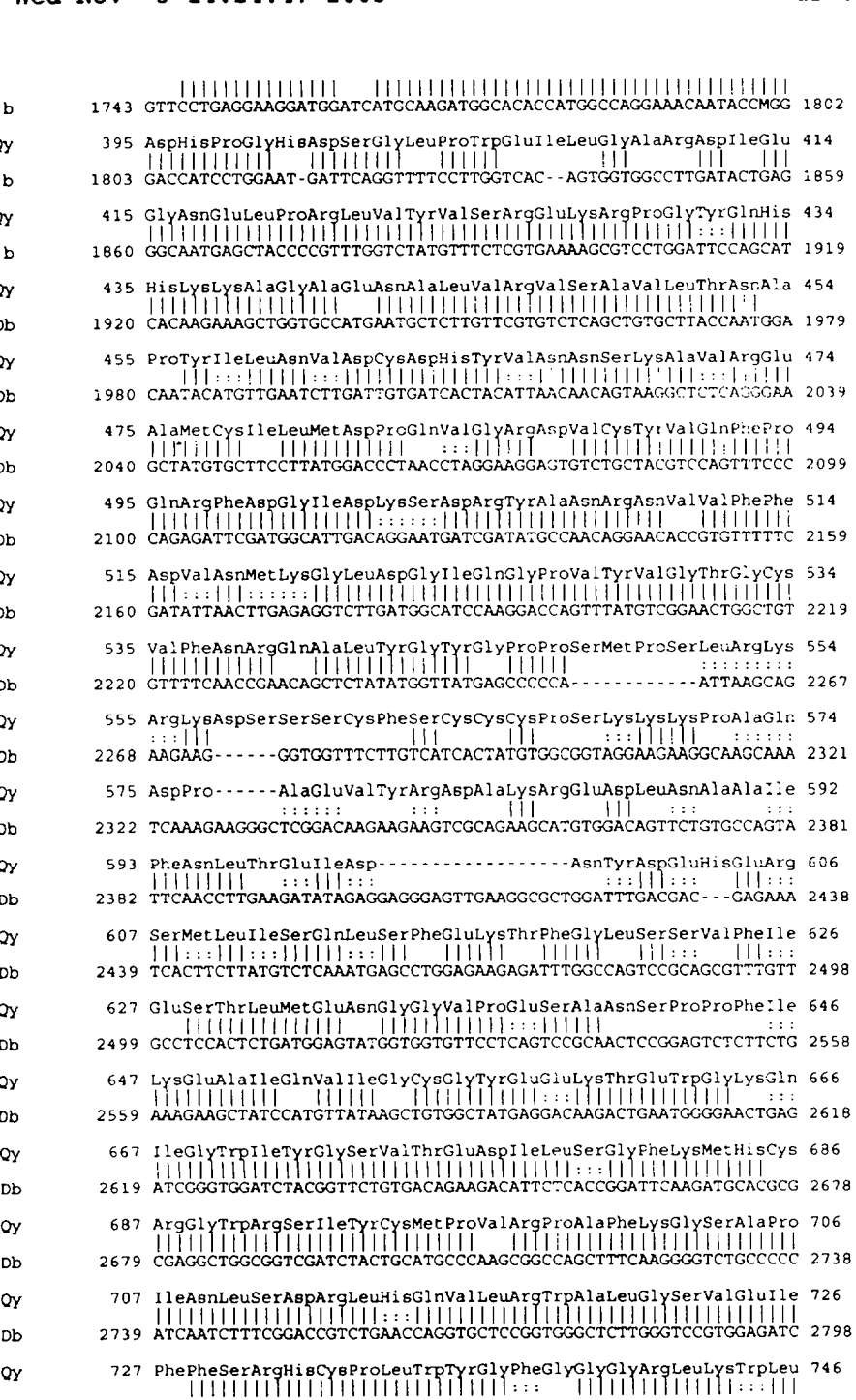
Alignment Scores:

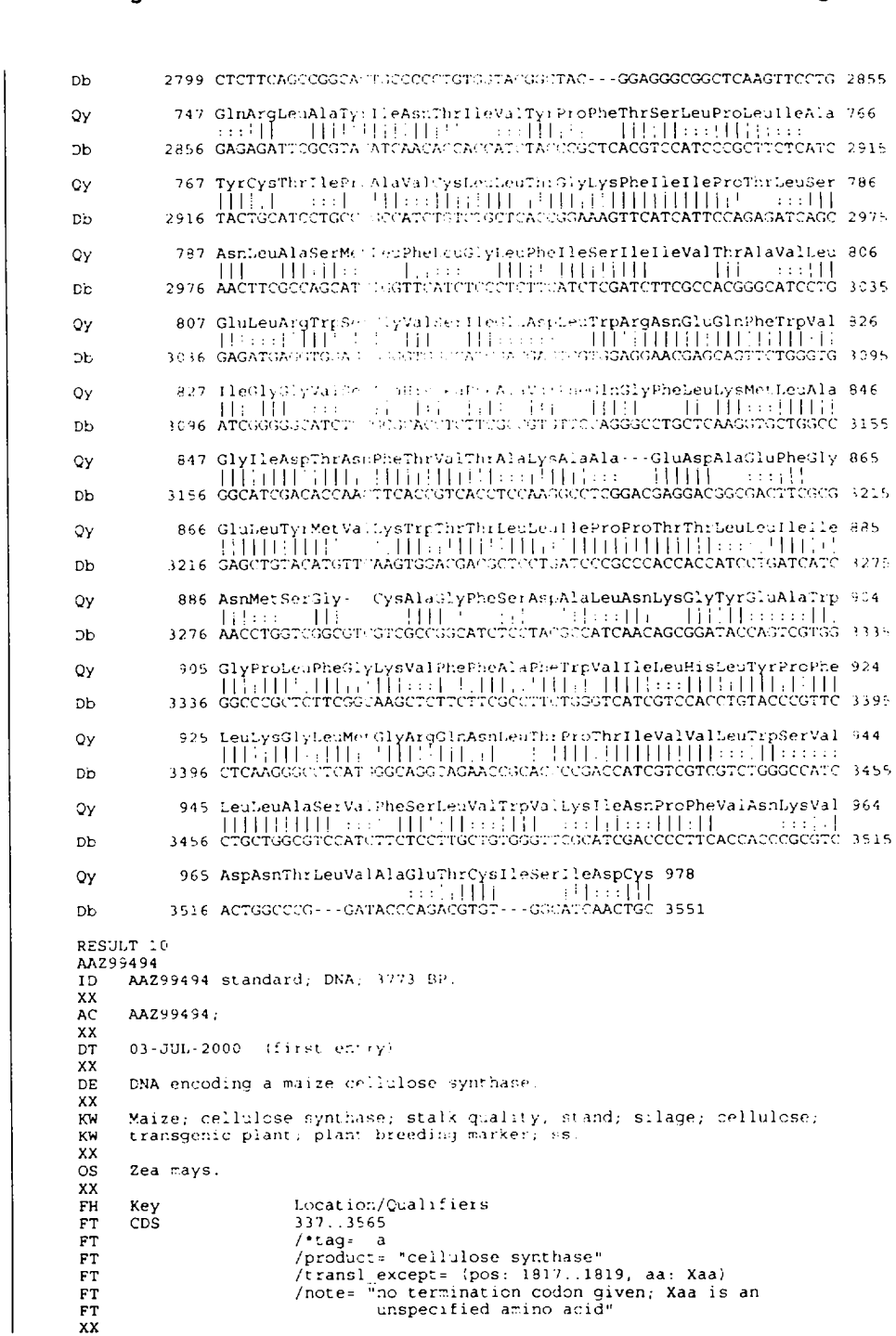
Pred. No.:	0	Length:	3746
Score:	3365.50	Matches:	648
Percent Similarity:	73.00%	Conservative:	196
Best Local Similarity:	60.34%	Mismatches:	140
Query Match:	64.23%	Indels:	17
DB:	21	Gaps:	3

US-09-980-043-2 (1-978) x AAZ99527 (1-3746)









RESULT 10
AAZ99494
ID AAZ99494 standard; DNA; 3773 BP.
XX
AC AAZ99494;
XX
DT 03-JUL-2000 (first entry)
XX
DE DNA encoding a maize cellulose synthase.
XX
KW Maize; cellulose synthase; stalk quality, stand; silage; cellulose;
KW transgenic plant; plant breeding marker; ss.
XX
OS Zea mays.
XX
FH Key Location/Qualifiers
FT CDS 337..3565
FT /*tag= a
FT /product= "cellulose synthase"
FT /transl_except= (pos: 1817..1819, aa: Xaa)
FT /note= "no termination codon given; Xaa is an
FT unspecified amino acid"
XX

PN WO200009706-A2.
XX
PD 24-FEB-2000.
XX
PF 16-AUG-1999; 99WO-US18760.
XX
PR 17-AUG-1998; 98US-0096822.
XX
PA (PION-) PIONEER HI-BRED INT INC.
XX
PI Dhugga KS, Helentjaris TG, Bowen BA, Wang X;
XX
DR WPI; 2000-224343/19.
DR P-PSDB; AAY84108.
XX
PT New genes which encode maize cellulose synthase polypeptides in plants
PT useful for modulating the expression of cellulose synthase in plants
PT and to produce transgenic plants expressing the novel protein -
XX
PS Claim 1; Page 90-94; 119pp; English.
XX
CC The present sequence encodes a maize cellulose synthase polypeptide.
CC The cellulose synthase can be used for the improvement of stalk quality
CC for improved stand or silage. It also provides an increased concentration
CC of cellulose in the pericarp, hardening the kernel and improving its
CC handling ability. The sequences are used to produce transgenic plants
CC and seeds expressing the cellulose synthase. The polynucleotide is
CC used for modulating, preferably increasing, the level of the synthase
CC in a plant cell. The plants are preferably monocots. The polynucleotide
CC is also used as a probe or primer in the detection quantitation or
CC isolation of gene transcripts. The probes are useful in detecting
CC deficiencies in the level of mRNA in screenings for desired transgenic
CC plant, for detecting mutations in the gene, for monitoring upregulation
CC of expression or changes in enzyme activity in screening assays of
CC compounds, for detection of any number of allelic variants of the gene,
CC or for use as molecular markers in plant breeding programs. The
CC isolated nucleic acids of the present invention can also be used for
CC recombinant expression of their encoded polypeptides or for use as
CC immunogens in the preparation and/or screening of antibodies. The
CC proteins can be employed in assays for enzyme agonists or antagonists
CC of enzyme function or for use of immunogens or antigens to obtain
CC antibodies specifically immunoreactive with a protein.
XX
SQ Sequence 3773 BP; 863 A; 988 C; 1017 G; 904 T; 1 other;

Alignment Scores:
Pred. No.: C Length: 3773
Score: 3365.50 Matches: 648
Percent Similarity: 73.00% Conservative: 136
Best Local Similarity: 60.34% Mismatches: 120
Query Match: 64.23% Indels: 4
DB: 21 Gaps: 2

US-09-980-043-2 (1-978) x AAZ99444 (1-3773)

```
Qy          5 GlyAlaProIleCysHisThrCysGlyGluGlnValGlyHisAspAlaAsnGlyGluLeu 24
              |||      :::|||     |||||:::   ||||     |||   |||::::::
Db        380 GGCGGACAGGTGTGCCAGATCTGCGGCGACGGCGTGGGCACCACGGCGGAGGGGGACGTC 439

Qy         25 PheValAlaCysHisGluCysSerTyrProMetCysLysSerCysPheGluPheGluIle 44
              |||   ||||||      |||   :::|||:::|:::   |||:::|:::|||
Db        440 TTCGCCGCCTGCGACGTCTGCGGGTTTCCGGTGTGCCGCCCCTGCTACGAGTACGAGCGC 499

Qy         45 AsnGluGlyArgLysValCysLeuArgCys--------------------GlySer 56
                :::|||   :::   :||   :::|||                    ||||||
Db        500 AAGGACGGCACGCAGGCGTGCCCCCAGTGCAAGACCAAGTACAAGCGCCACAAGGGGAGC 559

Qy         57 Pro------------------------------------TyrAspGluAsnLeuLeu 63
              |||                                  |||   |||   |||
Db        560 CCGGCGATCCGTGGGGAGGAAGGAGACGACACTGATGCCGATAGCGACTTCAATTACCTT 619

Qy         64 ------------AspAspValGluLysLysGlySerGlyAsnGlnSerThrMetAlaSer 79
                         :::|||   :::::::||   :::           :::

Db        620 GCATCTGGCAATGAGGACCAGAAGCAGAAGATTGCCGACAGAATGCGCAGCTGGCGCATG 679

Qy         80 HisLeuAsnAspSerGlnAspValGly---------------------------------- 88
              ::::::      |||   |||||||||
Db        680 AACGTTGGGGGCAGCGGGGATGTTGGTCGCCCCAAGTATGACAGTGGCGAGATCGGGCTT 739

Qy         89 ----------------------IleHisAlaArgHisIleSerSerValSer--------- 98
                                   |||         :::|||   ||||||:::
Db        740 ACCAAGTATGACAGTGGCGAGATTCCTCGGGGATACATCCCATCAGTCACTAACAGCCAG 799

Qy         98 ------------------------------------------------------------ 98

Db        800 ATCTCAGGAGAAATCCCTGGTGCTTCCCCTGACCATCATATGATGTCCCCAACTGGGAAC 859

Qy         99 ----------------------------------------ThrValAspSerGluMet 104
                                                      |||
Db        860 ATTGGCAAGCGTGCTCCATTTCCCTATGTGAACCATTCGCCAAATCCGTCAAGGGAGTTC 919

Qy        105 AsnAspGluTyrGlyAsnProIleTrpLysAsnArgValLysSerCysLysAspLysGlu 124
              :::         ||||||      ||||||   ||||||         |||   |||:::
Db        920 TCTGGTAGCATTGGGAATGTTGCCTGGAAAGAGAGGGTTGATGGCTGGAAAATGAAGCAG 979

Qy        125 AsnLys-----------LysLysLysArgSerProLysAlaGluThrGluProAlaGln 140
              :::|||                :::         |||   :::|||
Db        980 GACAAGGGGACGATTCCCATGACGAATGGCACAAGCATTGCTCCCTCTGAGGGTCGGGGT 1039

Qy        141 Val--------------ProThrGluGlnGlnMetGluGluLys---ProSerAlaGlu 154
              |||                  |||:::      ||||||:::         :::   |||
Db       1040 GTTGGTGATATTGATGCATCAACTGATTACAACATGGAAGATGCCTTATTGAACGACGAA 1099

Qy        155 AlaSerGluProLeuSerIleValTyrProIleProArgAsnLysLeuThrProTyrArg 174
                 :::|||||||||         |||:::|||   ::::::::   |||||||||
Db       1100 ACTCGACAGCCTCTATCTAGGAAAGTTCCACTTCCTTCCTCCAGGATAAATCCATACAGG 1159

Qy        175 AlaValIleIleMetArgLeuValIleLeuGlyLeuPhePheHisPheArgIleThrAsn 194
                 ||||||::::::|||||::::::|||   :::|||   |||:::|||||||||||||
Db       1160 ATGGTCATTGTGCTGCGATTGATTGTTCTAAGCATCTTCTTGCACTACCGTATCACAAAT 1219

Qy        195 ProValAspSerAlaPheGlyLeuTrpLeuThrSerValIleCysGluIleTrpPheAla 214
              ||||||   :::|||:::   |||||||||   ||||||||||||||||||||||||||||
Db       1220 CCTGTGCGCAATGCATACCCATTATGGCTTCTATCTGTTATATGTGAGATCTGGTTTGCT 1279

Qy        215 PheSerTrpValLeuAspGlnPheProLysTrpAsnProValAsnArgGluThrTyrIle 234
              ||||||:::|||||||||||||||||||||||   |||:::|||||||||||||||:::
Db       1280 CTTTCGTGGATATTGGATCAGTTCCCTAAGTGGTTTCCAATCAACCGGGAGACGTACCTT 1339

Qy        235 GluArgLeuSerAlaArgTyrGluArgGluGlyGluProSerGlnLeuAlaGlyValAsp 254
              :::||||||:::   ||||||:::|||||||||||||||||||||||||   ||||||
Db       1340 GATAGGCTGGCATTAAGGTATGACCGGGAAGGTGAGCCATCTCAGTTGGCTGCTGTTGAC 1399

Qy        255 PhePheValSerThrValAspProLeuLysGluProProLeuIleThrAlaAsnThrVal 274
              |||||||||||||||||:::|||||||||||||||||:::|||||||||||||||||
Db       1400 ATTTTCGTCAGTACAGTCGACCCAATGAAGGAGCCTCCTCTTGTCACTGCCAATACCGTG 1459

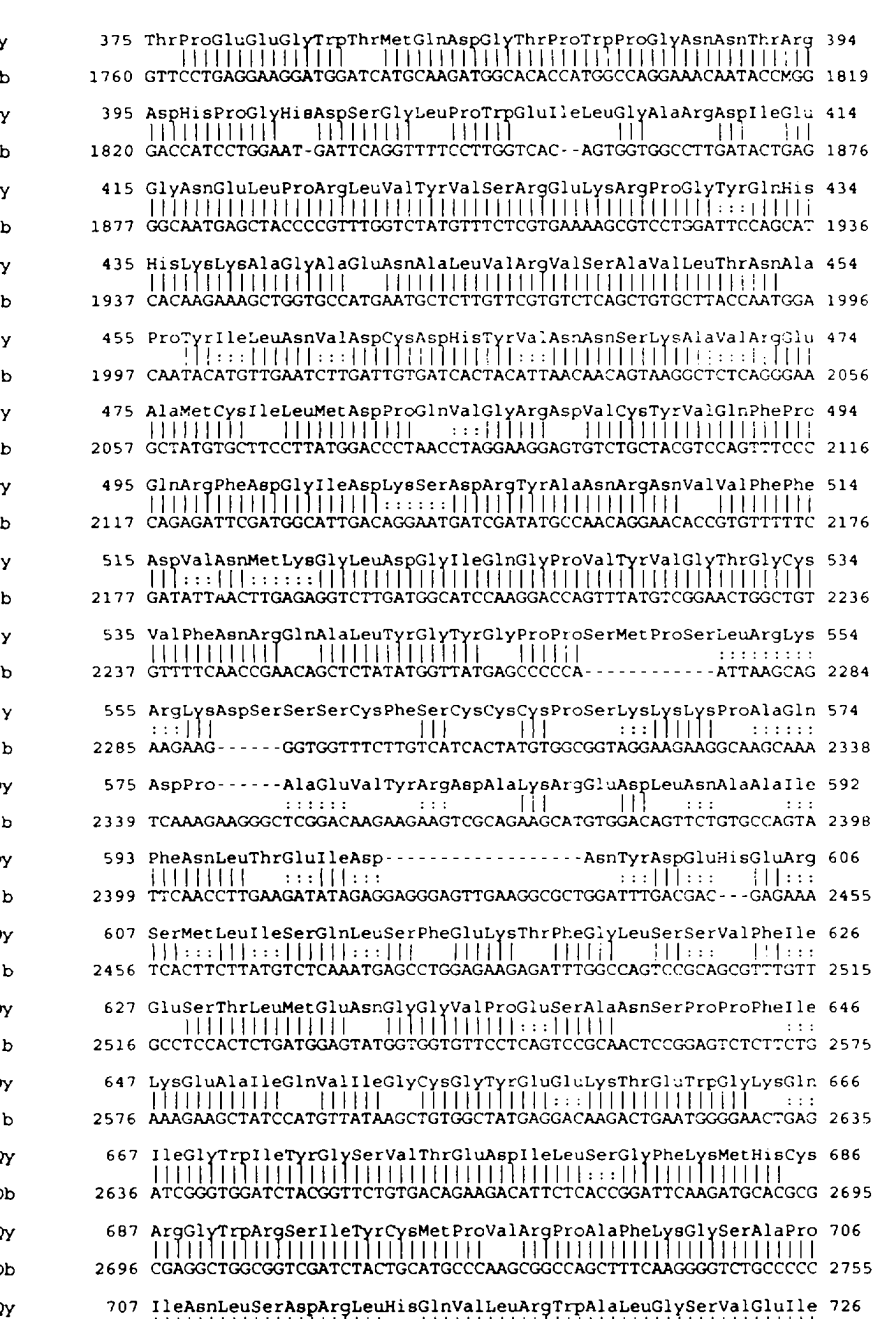
Qy        275 LeuSerIleLeuAlaValAspTyrProValAspLysValSerCysTyrValSerAspAsp 294
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1460 CTATCCATTCTTGCTGTGGATTACCCTGTGGATAAGGTCTCTTGCTATGTATCTGATGAT 1519

Qy        295 GlyAlaAlaMetLeuSerPheGluSerLeuValGluThrAlaGluPheAlaArgLysTrp 314
              ||||||||||||||:::|||::::::|||   ||||||:::||||||||||||||||||
Db       1520 GGAGCTGCGATGCTGACATTTGATGCACTAGCTGAGACTTCAGAGTTTGCTAGAAAATGG 1579

Qy        315 ValProPheCysLysLysPheSerIleGluProArgAlaProGluPheTyrPheSerGln 334
              |||||||||   ||||||:::::|||||||||||||||||||:::||||||||||||
Db       1580 GTACCATTTGTTAAGAAGTACAACATTGAACCTAGAGCTCCTGAATGGTACTTCTCCCAG 1639

Qy        335 LysIleAspTyrLeuLysAspLysValGlnProSerPheValLysGluArgArgAlaMet 354
              ||||||||||||||||||||||||   |||||||||||||:::||||||||||||
Db       1640 AAAATTGATTACTTGAAGGACAAAGTGCACCCTTCATTTGTTAAAGACCGCCGGGCCATG 1699

Qy        355 LysArgAspTyrGluGluTyrLysValArgValAsnAlaLeuValAlaLysAlaGlnLys 374
              ||||||:::|||||||||:::|||||||||||||||   ||||||||||||||||||||
Db       1700 AAGAGAGAATATGAAGAATTCAAAGTTAGGGTAAATGGCCTTGTTGCTAAGGCACAGAAA 1759
```





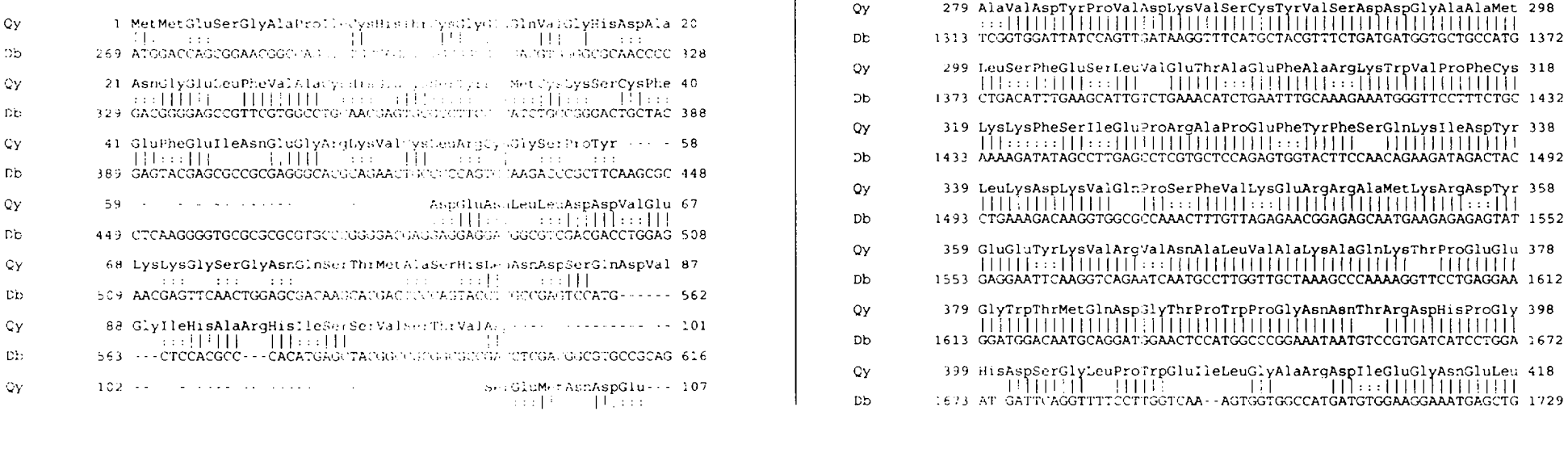
RESULT 11
AAZ58263
ID AAZ58263 standard; cDNA. 3776 BP.
XX
AC AAZ58263;
XX
DT 08-MAY-2000 (first entry)
XX
DE Corn cellulose synthase cDNA.
XX
KW Corn; maize; cellulose synthase; transgenic plant; ss.
XX
OS Zea mays.
XX
FH Key Location/Qualifiers
FT CDS 2..3448
FT /*tag= a
FT /partial
XX
PN WO200004166-A2.
XX

PD 27-JAN-2000.
XX
PF 13-JUL-1999; 99WO-US15871.
XX
PR 14-JUL-1998; 98US-0092844.
XX
PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX
PI Allen SM, Fader GM, Falco SC, Kinney AJ, Lightner JE, Miao G;
PI Rafalski JA, Thorpe CJ;
XX
DR WPI; 2000-182431/16.
DR P-PSDB; AAY58832.
XX
PT Nucleic acid fragments encoding cellulose biosynthetic enzyme useful as
PT probes for isolating cDNAs and genes encoding homologous proteins, for
PT producing transgenic plants -
XX
PS Claim 9; Page 37-38; 93pp; English.
XX
CC The present sequence is that of a contig of cDNA clones and encodes
CC a portion (see AAY58832) of corn cellulose synthase (CS). The cDNA
CC clones were isolated from corn cob, developing tassel, seedling,
CC shoot culture and leaf tissue cDNA libraries on the basis of
CC homology to Arabidopsis and cotton CS sequences. The invention
CC relates to isolated nucleic acid fragments encoding plant CS and to
CC CS polypeptides. It also relates to the construction of a chimeric
CC gene encoding all or a portion of the CS, in sense or antisense
CC orientation, where expression of the gene results in altered levels
CC of the CS in transformed host cells. The host cells can be used to
CC screen compounds for their ability to inhibit CS activity. CS
CC nucleic acids are also useful for producing transgenic plants having
CC altered levels of CS, and hence altered levels of fibre. CS may
CC also serve as a target for the development of novel herbicides.
XX
SQ Sequence 3776 BP; 931 A; 906 C; 999 G; 940 T; 0 other;

Alignment Scores:

Pred. No.:	0	Length:	3776
Score:	3363.00	Matches:	632
Percent Similarity:	72.73%	Conservative:	152
Best Local Similarity:	58.63%	Mismatches:	175
Query Match:	64.18%	Indels:	121
DB	21	Gaps:	15

US-09-980-043-2 (1-978) x AAZ58263 (1-3776)



```
y     419 ProArgLeuValTyrValSerArgGluLysArgProGlyTyrGlnHisHisLysLysAla 438
b    1730 CCTCGATTGGTTTATGTTTCAAGAGAAAAACGGCCAGGCTACAACCATCACAAGAAGGCT 1789

y     439 GlyAlaGluAsnAlaLeuValArgValSerAlaValLeuThrAsnAlaProTyrIleLeu 458
b    1790 GGTGCTATGAATGCATTGGTCCGAGTCTCTGCTGTACTAACTAATGCTCCTTATTTGCTG 1849

y     459 AsnValAspCysAspHisTyrValAsnAsnSerLysAlaValArgGluAlaMetCysIle 478
b    1850 AACTTGGATTGTGATCACTATATCAATAATAGTAAGGCTATAAAGGAAGCAATGTGTTTT 1909

y     479 LeuMetAspProGlnValGlyArgAspValCysTyrValGlnPheProGlnArgPheAsp 498
b    1910 ATGATGGATCCTTTGCTTGGAAAGAAAGTTTGCTATGTGCAGTTTCCTCAAAGATTTGAT 1969

y     499 GlyIleAspLysSerAspArgTyrAlaAsnArgAsnValValPhePheAspValAsnMet 518
b    1970 GGGATTGATCGCCATGATCGATATGCTAACAGAAATGTTGTCTTTTTCGATATCAACATG 2029

y     519 LysGlyLeuAspGlyIleGlnGlyProValTyrValGlyThrGlyCysValPheAsnArg 538
b    2030 AAAGGTTTGGATGGTATCCAGGGCCCAATTTATGTGGGTACTGGATGTGTCTTCAGAAGG 2089

y     539 GlnAlaLeuTyrGlyTyrGlyProProSerMetProSerLeuArgLysArgLysAspSer 558
b    2090 CAGGCATTATATGGCTACGATGCTCCCAAA------- ----ACAAAGAAGCCACCATCA 2137

y     559 SerSer--------------------CysPheSerCysCysCys----- --------- 566
b    2138 AGAACTTGCAACTGCTGGCCAAAGTGGTGCATTTGCTGTTGCTGTTTTGGTAACAGGAAG 2197

y     567 ProSerLysLysLysProAlaGlnAspPro------AlaGluValTyrArgAspAlaLys 584
b    2198 ACCAAGAAGAAGACCAAGACCTCTAAACCTAAATTTGAGAAGATAAAGAAACTTTTTAAG 2257

y     585 ArgGluAspLeuAsnAlaAlaIlePheAsnLeuThrGluIleAspAsnTyr--------- 601
b    2258 AAAAAGGAAAATCAAGCCCCTGCATATGCTCTTGGTGAAATTGATGAAGCCGCTCCAGGA 2317

y     602 AspGluHisGluArgSerMetLeuIleSerGlnLeuSerPheGluLysThrPheGlyLeu 621
b    2318 GCTGAAAATGAAAAGGCTAGTATTGTAAATCAACAGAAGTTGGAAAAGAAATTTGGCCAG 2377

y     622 SerSerValPheIleGluSerThrLeuMetGluAsnGlyGlyValProGluSerAlaAsn 641
b    2378 TCTTCAGTTTTTGTTGCATCCACACTTCTTGAGAATGGTGGAACCCTGAAGAGTGCCAGT 2437

y     642 SerProProPheIleLysGluAlaIleGlnValIleGlyCysGlyTyrGluGluLysThr 661
b    2438 CCAGCTTCTCTTCTGAAGGAAGCTATACATGTCATCAGTTGTGGATATGAAGACAAAACA 2497

y     662 GluTrpGlyLysGlnIleGlyTrpIleTyrGlySerValThrGluAspIleLeuSerGly 681
b    2498 GGCTGGGGAAAAGATATTGGTTGGATTTATGGATCAGTCACAGAAGATATTCTTACTGGG 2557

y     682 PheLysMetHisCysArgGlyTrpArgSerIleTyrCysMetProValArgProAlaPhe 701
b    2558 TTTAAGATGCACTGCCATGGTTGGCGGTCAATTTACTGCATACCTAAACGGGCCGCCTTC 2617

y     702 LysGlySerAlaProIleAsnLeuSerAspArgLeuHisGlnValLeuArgTrpAlaLeu 721
b    2618 AAAGGTTCCGCACCTCTCAATCTTTCCGATCGTCTTCACCAGGTTCTTCGGTGGGCTCTT 2677

y     722 GlySerValGluIlePhePheSerArgHisCysProLeuTrpTyrGlyPheGlyGlyGly 741
b    2678 GGTTCAATTGAAATTTTCTTCAGCAACCACTGCCCTCTCTGGTATGGGTATGGTGGTGGA 2737

y     742 ArgLeuLysTrpLeuGlnArgLeuAlaTyrIleAsnThrIleValTyrProPheThrSer 761
b    2738 ---CTAAAGTTCCTGGAAAGGTTTTCGTACATTAACTCCATCGTATACCCTTGGACATCT 2794
```

```
Qy     762 LeuProLeuIleA  TyrCysTh  ProAlaValCysLeuLeuThrGlyLysPheIle 781
Db    2795 ATCCCGCTCTTGGC  ATTGCACA  TGCTTGC ATCTGCTTGCTGACAGGGAAATTTATC 2854

Qy     782 IleProThrLeuSe  AsnLeuAlaSe  MetLeuPheLeuGlyLeuPheIleSerIleIle 801
Db    2855 ACGCCAGAGCTTAA  AATGTCA  TCTTGC TTCATGTCACTTTTCATCTGCATTTTT 2914

Qy     802 ValThrAlaValIe  GluLeuArgTrpSerGlyValSerIleGluAspLeuTrpArgAsn 821
Db    2915 GCTACGAGCATCCT  AAATGAGATGGAGTGGTGTAGGCATCGATGACTGGTGGAGAAAC 2974

Qy     822 GluGlnPheTrpVa  IleGlyGlyValSerAlaHisLeuPheAlaValPheGlnGlyPhe 841
Db    2975 GAGCAGTTTTGGGT  ATTGGA  GGTCTTCACATCTCTTTGCTGTGTTCCAGGGACTC 3034

Qy     842 LeuLysMe  A  lyr  A  g  As PheThrValThrAlaLysAlaAlaGluAsp 861
Db    3035 CTCAAG  TA  A  A  A  A  A  GACATCCAAGGGCA AGACGAC 3094

Qy     862 AlaGluPr  Gly  euLysMe  Va  ysTr  hrThrLeuLeuIleProProThrThr 881
Db    3095 GAGGAG  TCAGA  TGTACACATTCAAATG  ACGACCTTCTGATACCTCCGACAACC 3154

Qy     882 LeuLeuIleIleAsnMetSerGly---CysA  GlyPheSerAspAlaLeuAsnLysGly 900
Db    3155 CTGCTCCTATTGAA TTCATTGGAG GGTAG  GGCATCTCCAATGCGATCAACAAGGGA 3214

Qy     901 TyrGluAlaTrpGlyProLeuPheGlyLysVa  PhePheAlaPheTrpValIleLeuHis 920
Db    3215 TATGAATCATGGGG  CCCTGTTTGG GAAGCT  TTCTTTGCATTTTGGGTGATCGTCCAT 3274

Qy     921 LeuTyrProPheLe  LysGlyLeuMetGlyAr  GlnAsnLeuThrProThrIleValVal 940
Db    3275 CTTTACC  TTCCT  AAGGGTCTGATTGGGAGG  AGAACAGGACGCCAACGATTGTCATT 3334

Qy     941 LeuTrpSerValLe  LeuAlaSe  ValPheSe  LeuValTrpValLysIleAsnProPhe 960
Db    3335 GTCTGGTC  CT  CT  CTGGCTTCGATCTTCTCGCTGCTTTGGGTCCGGATCGACCCGTTC 3394

Qy     961 ValAsnLysValAspAsnThrLeuValAlaGlnThrCysIleSerIleAspCys 978
Db    3395 CTTGCGAAGGATGA  GGTCCGCTGTTGGAGGA  --- --TGTGGTCTGGATTGC 3442
```

```
RESULT 12
AAZ58268
ID   AAZ58268 standard; cDNA; 3517 BP.
XX
AC   AAZ58268;
XX
DT   08-MAY-2000  (first entry)
XX
DE   Soybean cellulose synthase cDNA clone sdp:pk005.o22.
XX
KW   Soybean; cellulose synthase; transgenic plant; ss.
XX
OS   Glycine max.
XX
FH   Key             Location/Qualifiers
FT   CDS             158..32 8
FT                   /*tag= a
FT                   /transl_except= (pos:758..761, aa:Xaa)
FT                   /note= "apparent 1 base addition, alters the
FT                   reading frame"
XX
PN   WO200004166-A2.
XX
PD   27-JAN-2000.
XX
PF   13-JUL-1999;   99WO-US15871.
XX
PR   14-JUL-1998;   98US-0092844.
XX
```

```
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Allen SM,  Fader GM,  Falco SC,  Kinney AJ,  Lightner JE,  Miao G;
PI   Rafalski JA,  Thorpe CJ;
XX
DR   WPI; 2003-182431/16.
DR   P-PSDB; AAY58837.
XX
PT   Nucleic acid fragments encoding cellulose biosynthetic enzyme useful as
PT   probes for isolating cDNAs and genes encoding homologous proteins, for
PT   producing transgenic plants
XX
PS   Claim 3; Page 55-56; 93pp; English.
XX
CC   The present sequence is that of full-length cDNA clone
CC   sdp2c.pk005.o22 encoding soybean cellulose synthase (CS, see AAY58837).
CC   The cDNA clone was isolated from a cDNA library produced from
CC   soybean pod tissue.  The invention relates to isolated nucleic acid
CC   fragments encoding plant CS and to CS polypeptides.  It also relates
CC   to the construction of a chimeric gene encoding all or a portion of
CC   the CS, in sense or antisense orientation, where expression of the
CC   gene results in altered levels of the CS in transformed host cells.
CC   The host cells can be used to screen compounds for their ability to
CC   inhibit CS activity.  CS nucleic acids are also useful for producing
CC   transgenic plants having altered levels of CS, and hence altered
CC   levels of fibre.  CS may also serve as a target for the development
CC   of novel herbicides.
XX
SQ   Sequence 3517 BP; 978 A; 686 C; 841 G; 1012 T, 0 other;

Alignment Scores.
Pred. No.:              0           Length:         3517
Score:                  3362.00     Matches:        1146
Percent Similarity:     74.42%      Conservative:    119
Best Local Similarity:  62.84%      Mismatches:      179
Query Match:            64.16%      Indels:           86
DB:                     21          Gaps:             15

US-09-980-043-2 (1-978) x AAZ58268 (1-3517)

Qy          8 IleCysHisThrCysGlyGluGlnValGlyHisAspAlaAsnGlyGluLeuPheValAla 27
              :::|||     |||||:::   |'|||       :::|||:::|||||||||||
Db        263 GTGTGTGAGATTTGTGGTGATGGCGTGGGACTCACGGTGGATGGAGACTTGTTTGTGGCT 322

Qy         28 CysHisGluCysSerTyrProMetCysLysSerCysPheGluPheGluIleAsnGluGly 47
              |||:::|||||   :::|||:::|||:::   |||:::|||:::|||      ||||||
Db        323 TGCAATGAGTGTGGTTTTCCAGTGTGCAGGCCTTGCTATGATGTATGAAAGGAGAGAAGGA 382

Qy         48 ArgLysValCysLeuArgTyr                        GlySerProTyr  58
                 :::||    :::|                           |:  |||:|
Db        383 AGCCAACTTTGCCCACAGTGTAAAAGTAGATAGAAGTGTTAAAGTGAGCCCCCGAGTG 442

Qy         59 -- ---------AspGluAsnLeuLeuAspAspValGlnLysLys- --- --------- 69
                       |||||     :::.|||  :::||  :::
Db        443 GAGGGAGATGATGATGAAGAGGATGTGGATGATATTGAGCATGAATTCAATATTGATGAG 502

Qy         70      ----------------------------------------     -------Gly 70
                                                                      |||
Db        503 CAAAAGAACAAGCATGGCCAGGTTGCAGAAGCCATGCTTCATGGGAGGATGAGCTATGGA 562

Qy         71 SerGlyAsnGlnSerThrMetAlaSerHisLeuAsnAspSerGlnAspValGlyIleHis 90
                 |||   :::             |'|                    :||
Db        563 AGAGGTCCTGAAGATGATGACAATTGCCAGTTCCCAACACATGTCATTGCTGGTGGTCGT 622

Qy         91 AlaArgHisIleSerSerValSerThrValAspSerGluMetAsnAspGluTyrGlyAsn 110
              :::|||   :::|||          :::   |||         :::   |:||||:::
Db        623 TCTAGGCCTGTAAGTGGGGAGTTCCCAATATCATCT---    ---AATGCTTATGGGGAT 673

Qy        111 ProIle-TrpLys---AsnArgValLysSerCys------  ------------------ 120
                 ::: :::      :::   :::||   | |
Db        674 CAGATGTTATCCTCTTCACTGCATAAAAGAGTGCATCCATATCCAGTGTCTGAACCTGGA 733

Qy        121  - -------------LysAspLysGluAsnLysLysLysLysArgSerProLysAlaGl 135

                          |||   ||||||:::      |||   |||            ::
Db        734 AGTGCAAGATGGGACGAAAAAAAAAGAAGATGGATGGAAAGATAGAATGGATGACTGGAA 793

Qy        135 uThrGluProAlaGlnValProThrGluGlnGlnMetGluGluLysProSerAla----- 153
              :  :::         :::      |||         :::|||   |||   |||
Db        794 ATTGCAGCAAGGCAATTTGGGGCCTGAACCG------GATGAAGATCCAGATGCAGCCAT 847

Qy        154 -------GluAlaSerGluProLeuSerIleValTyrProIleProArgAsnLysLeuTh 171
                    ||||||   :::|||||||||          ||||||      :::|||:::
Db        848 GTTAGATGAAGCAAGGCAACCACTGTCAAGGAAAGTGCCAATAGCATCCAGCAAAATCAA 907

Qy        171 rProTyrArgAlaValIleIleMetArgLeuValIleLeuGlyLeuPhePheHisPheAr 191
              ||||||||   |||||||:::  |||||||||||||||   |||      :::||
Db        908 TCCATATAGAATGGTGATTGTGGCACGTCTGGTTATTCTTGCTTTCTTCCTCAGATACAG 967

Qy        191 gIleThrAsnProValAspSerAlaPheGlyLeuTrpLeuThrSerValIleCysGluIl 211
              |:::   |||||||||      |||   |||||||||||||||||:::||||||||||
Db        968 ACTCATGAACCCAGTACATGATGCCCTGGGGCTATGGCTAACCTCTATCATATGTGAAAT 1027

Qy        211 eTrpPheAlaPheSerTrpValLeuAspGlnPheProLysTrpAsnProValAsnArgGl 231
              ||||||||||||||||||:::|||||||||||||||||||||    |||:::::|||||
Db       1028 CTGGTTTGCTTTTTCATGGATTCTGGATCAGTTTCCCAAATGGTTTCCCATTGATAGAGA 1087

Qy        231 uThrTyrIleGluArgLeuSerAlaArgTyrGluArgGluGlyGluProSerGlnLeuAl 251
              |||||||::::::|||||||||   ||||||||||||||||||||||||:::   |||||
Db       1088 GACCTACCTTGACCGTCTTTCCATCAGGTATGAGCGTGAAGGTGAACCAAACATGCTTGC 1147

Qy        251 aGlyValAspPhePheValSerThrValAspProLeuLysGluProProLeuIleThrAl 271
              |   ||||||   ||||||||||||||||||||||:::||||||||||||||:::|||||
Db       1148 TCCTGTAGATGTTTTTGTTAGTACCGTGGATCCCATGAAGGAACCTCCTCTGGTTACAGC 1207

Qy        271 aAsnThrValLeuSerIleLeuAlaValAspTyrProValAspLysValSerCysTyrVa 291
              ||||||||||||||||||||||||||:::|||||||||||||||||:::|||||||::
Db       1208 AAACACTGTTCTTTCAATCTTGGCCATGGATTACCCGGTTGATAAAATATCATGCTACAT 1267

Qy        291 lSerAspAspGlyAlaAlaMetLeuSerPheGluSerLeuValGluThrAlaGluPheAl 311
              :|||||||||||||||:::|||   :::|||||||||||||   ||||||||||||||||
Db       1268 TTCTGATGATGGAGCCTCAATGTGTACATTTGAGTCCTTATCAGAAACTGCAGAGTTTGC 1327

Qy        311 aArgLysTrpValProPheCysLysLysPheSerIleGluProArgAlaProGluPheTy 331
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||   ||
Db       1328 TAGAAAGTGGGTACCGTTTTGTAAGAAATTTTCCATAGAACCTCGGGCACCTGAGATGTA 1387

Qy        331 rPheSerGlnLysIleAspTyrLeuLysAspLysValGlnProSerPheValLysGluAr 351
              |||||||:::|||||||||||||||||||||||||||||||:::|||||||||||||||
Db       1388 CTTCAGCGAGAAGATTGACTACCTAAAGGACAAAGTGCAACCCACCTTTGTTAAGGAGCG 1447

Qy        351 gArgAlaMetLysArgAspTyrGluGluTyrLysValArgValAsnAlaLeuValAlaLy 371
              |||||||||||||||:::|||||||||:::|||||||||:::||||||||||||||||||
Db       1448 TCGAGCTATGAAGAGGGAATACGAAGAGTTTAAGGTTAGGATCAATGCACTTGTTGCTAA 1507

Qy        371 sAlaGlnLysThrProGluGluGlyTrpThrMetGlnAspGlyThrProTrpProGlyAs 391
              |||||||||   |||:::   ||||||   ||||||||||||||||||||||||||||||
Db       1508 GGCCCAGAAAGTTCCTCAGGGAGGATGGATCATGCAGGATGGGACACCATGGCCAGGGAA 1567

Qy        391 nAsnThrArgAspHisProGlyHisAspSerGlyLeuProTrpGluIleLeuGlyAlaAr 411
              |||||||:::|||||||||||   |||||||   :::   |||          |||
Db       1568 TAACACTAAGGATCATCCTGGTAT-GATTCAAGTGTTTCTTGGT--AGCAGTGGAGGTCT 1624

Qy        411 gAspIleGluGlyAsnGluLeuProArgLeuValTyrValSerArgGluLysArgProGl 431
              |||   ||||||||||:::||||||||||||||||||||||||||||||||||||||||
Db       1625 TGATACTGAAGGAAACCAACTTCCTCGCCTTGTTTATGTTTCCAGAGAGAAAAGGCCTGG 1684

Qy        431 yTyrGlnHisHisLysLysAlaGlyAlaGluAsnAlaLeuValArgValSerAlaValLe 451
              |:::|||||||||||||||||||||||   ||||||||||||||||||||||||||||||
Db       1685 TTTTCAACACCACAAGAAAGCTGGTGCCATGAATGCTCTGGTTCGGGTATCTGCTGTTCT 1744

Qy        451 uThrAsnAlaProTyrIleLeuAsnValAspCysAspHisTyrValAsnAsnSerLysAl 471
              |||||||||||||:::::|||||:::|||||||||||||||||||||||||||||||||
Db       1745 CACAAATGCTCCTTTCATGTTGAACTTGGATTGTGATCACTATGTCAATAACAGCAAGGC 1804

Qy        471 aValArgGluAlaMetCysIleLeuMetAspProGlnValGlyArgAspValCysTyrVa 491
              |  ||||||||||:|||||   ||||||||||||||   |||:::   ||||||||||
```

```
b     1805 TGCCCGAGAGGCCATGTGCTTCTTGATGGACCCACAAACTGGGAAGAAGGTCTGCTATGT 1864

y      491 lGlnPheProGlnArgPheAspGlyIleAspLysSerAspArgTyrAlaAsnArgAsnVa 511
b     1865 CCAGTTTCCTCAAAGATTTGATGGTATTGATACACATGATCGTTATGCCAACAGGAACAC 1924

y      511 lValPhePheAspValAsnMetLysGlyLeuAspGlyIleGlnGlyProValTyrValGl 531
b     1925 AGTTTCTTTGATATTAACATGAAGGGTCTAGATGGTATTCAAGGTCCTGTATATGTGGG 1984

y      531 yThrGlyCysValPheAsnArgGlnAlaLeuTyrGlyTyrGlyProProSerMetProSe 551
b     1985 GACTGGATGTGTTTTCAGGAGGCAAGCTTTGTATGGCTATAATCCTCCCAAGGGTCCA-- 2042

y      551 rLeuArgLysArgLysAspSerSerSerCysPheSerCys---CysCysProSerLysLy 570
b     2043 -------AAGCGTCCAAAAATGGTAAGCTGTGATTGTTGCCCGTGTTTGGAAGCCGCAA 2095

y      570 sLysProAlaGlnAspProAlaGluValTyrArgAspAlaLysArgGluAspLeuAsnAl 590
b     2096 GAAG-----------------------TATAAGGAG---- AAGAATGATGCAAATGG 2125

y      590 aAlaIlePheAsnLeuThrGluIleAspAsnTyrAspGluHisGluArgSerMetLeuIl 610
b     2126 AGAGGCTGCAAGCCTAAAAGGGATGGATGAT-----------GACAAAGAGGTGTTGAT 2173

y      610 eSerGlnLeuSerPheGluLysThrPheGlyLeuSerSerValPheIleGluSerThrLe 630
b     2174 GTCCCAAATGAATTTTGAGAAGAAATTTGGACAATCCTCTATTTTTGTGACTTCTACCTT 2233

y      630 uMetGluAsnGlyGlyValProGluSerAlaAsnSerProProPheIleLysGluAlaIl 650
b     2234 GATGGAAGAGGGTGGTGTGCCTCCTTCTTCAAGTCCAGCTGCCCTGCTTAAAGAAGCCAT 2293

y      650 eGlnValIleGlyCysGlyTyrGluGluLysThrGluTrpGlyLysGlnIleGlyTrpIl 670
b     2294 TCATGTGATTAGCTGTGGATATGAAGATAAAACTGAATGGGGACTTGAGCTTGGTTGGAT 2353

y      670 eTyrGlySerValThrGluAspIleLeuSerGlyPheLysMetHisCysArgGlyTrpAr 690
b     2354 CTATGGATCTATCACAGAAGATATTCTAACAGGTTTTAAGATGCATTGCCGTGGGTGGAG 2413

y      690 gSerIleTyrCysMetProValArgProAlaPheLysGlySerAlaProIleAsnLeuSe 710
b     2414 GTCCATTTATTGTATGCCAAAGAGAGCTGCATTCAAGGGTACTGCTCCTATCAACTTGTC 2473

y      710 rAspArgLeuHisGlnValLeuArgTrpAlaLeuGlySerValGluIlePhePheSerAr 730
b     2474 AGATCGTCTCAACCAGGTTCTTCGTTGGGCACTTGGTTCCATTGAGATTTTCTTTAGTCA 2533

y      730 gHisCysProLeuTrpTyrGlyPheGlyGlyGlyArgLeuLysTrpLeuGlnArgLeuAl 750
b     2534 CCATTGCCCTCTATGGTATGGCTTCAAGGAAAAGAAGCTAAAGTGGCTTGAGAGATTTGC 2593

y      750 aTyrIleAsnThrIleValTyrProPheThrSerLeuProLeuIleAlaTyrCysThrIl 770
b     2594 CTATGCAAACACAACTGTCTATCCATTCACCTCCATTCCTCTAGTTGCCTACTGTATTCT 2653

y      770 eProAlaValCysLeuLeuThrGlyLysPheIleIleProThrLeuSerAsnLeuAlaSe 790
b     2654 TCCAGCAGTTTGTTTACTCACTGACAAATTCATCATGCCACCGATTAGCACCTTTGCTGG 2713

y      790 rMetLeuPheLeuGlyLeuPheIleSerIleIleValThrAlaValLeuGluLeuArgTr 810
b     2714 TTTGTACTTTGTTGCTCTCTTCTCCTCAATCATTGCAACTGGTATTCTTGAGTTGAAATG 2773

y      810 pSerGlyValSerIleGluAspLeuTrpArgAsnGluGlnPheTrpValIleGlyGlyVa 830
b     2774 GAGTGGAGTGAGCATTGAGGAATGGTGGAGAAATGAGCAGTTTTGGGTCATTGGTGGTGT 2833

y      830 lSerAlaHisLeuPheAlaValPheGlnGlyPheLeuLysMetLeuAlaGlyIleAspTh 850
b     2834 ATCAGCTCACCTCTTTGCTGTTATACAAGGTCTGCTAAAGGTTCTGGCTGGAATTGACAC 2893

Qy     850 rAsnPheThrValThrAlaLysAlaAlaGluAspAlaGluPheGlyGluLeuTyrMetVa 870
Db    2894 CAATTTCACTGTTACATCAAAGGCAACAGATGATGAAGAGTTTGGAGAATTGTACACCTT 2953

Qy     870 lLysTrpThrThrLeuLeuIleProProThrThrLeuLeuIleIleAsnMetSerGly-- 889
Db    2954 TAAGTGGACTACACTTTTGATTCCTCCAACCACTATTTTGATCATTAACATTGTTGGTGT 3013

Qy     890 -CysAlaGlyPheSerAspAlaLeuAsnLysGlyTyrGluAlaTrpGlyProLeuPheGl 909
Db    3014 TGTTGCTGGAATCTCAGATGCTATAAACAATGGGTACCAATCCTGGGGACCACTCTTGG 3073

Qy     909 yLysValPhePheAlaPheTrpValIleLeuHisLeuTyrProPheLeuLysGlyLeuMe 929
Db    3074 AAAGCTCTTTTTT[illegible]ATCCATTCCTTAAAGGTTTGAT 3133

Qy     929 tGly[illegible]pSerValLeuLeuAlaSerVa 949
Db    3134 [illegible]GGTCAGTGCTATTG[illegible]CTAT 3193

Qy     949 lPheSerLeuValTrpValLysIleAsnProPheValAsnLysVal-- - --AspAs 966
Db    3194 TTTCTCCTTACTTTGGGTAAGAATTGATCCATTTGTCCTCAAGACTAAGGGACCTGATAC 3253

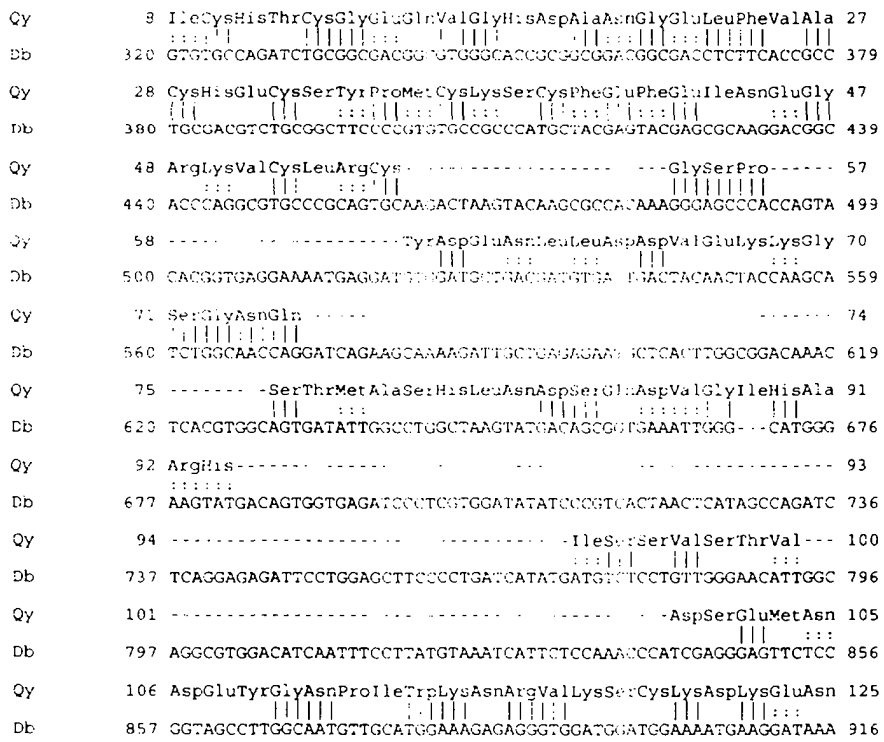
Qy     966 nThrLeuValAlaGluThrCys 973
Db    3254 CAAGCTATTGGAATTAACTGC 3275

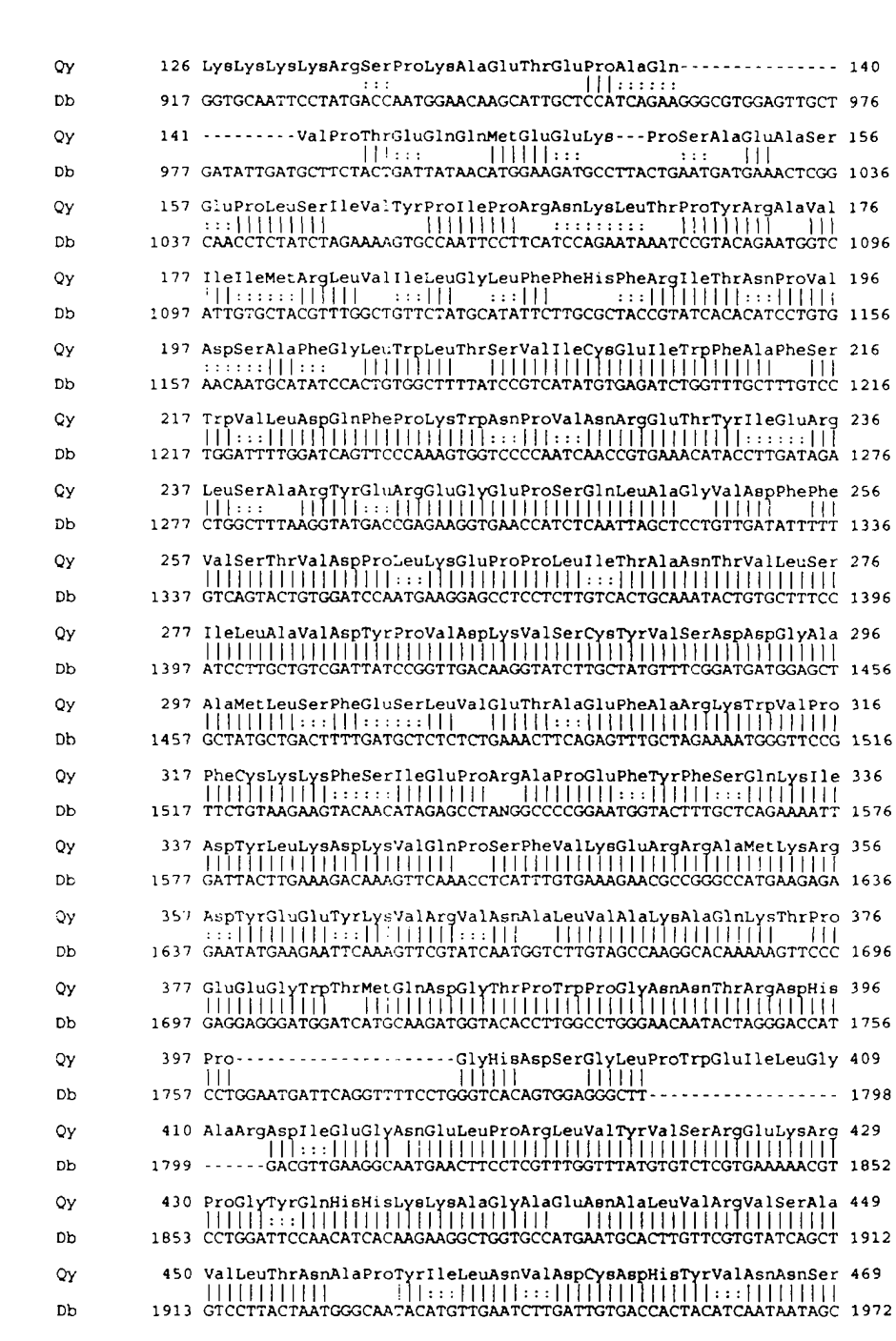
RESULT 13
AAZ99533
ID   AAZ99533 standard; DNA; 3704 BP.
XX
AC   AAZ99533;
XX
DT   03-JUL-2000  (first entry)
XX
DE   DNA encoding a maize cellulose synthase.
XX
KW   Maize; cellulose synthase; stalk quality; stand; silage; cellulose;
KW   transgenic plant; plant breeding marker; ss.
XX
OS   Zea mays.
XX
FH   Key             Location/Qualifiers
FT   CDS             272..3496
FT                   /*tag= a
FT                   /product= "cellulose synthase"
FT                   /transl except= (pos: 1544..1546, aa: Xaa)
FT                   /note= "no termination codon given; Xaa is an
FT                          unspecified amino acid"
XX
PN   WO200009706-A2.
XX
PD   24-FEB-2000.
XX
PF   16-AUG-1999;   99WO-US18760.
XX
PR   17-AUG-1998;   98US-0096822.
XX
PA   (PION-) PIONEER HI BRED INT INC.
XX
PI   Dhugga KS,  Helentjaris TG,  Bowen BA,  Wang X;
XX
DR   WPI; 2000-224343/19.
DR   P-PSDB; AAY84121.
XX
PT   New genes which encode maize cellulose synthase polypeptides in plants
PT   useful for modulating the expression of cellulose synthase in plants
PT   and to produce transgenic plants expressing the novel protein  -
XX
PS   Claim 1; Page 191-196; 119pp; English.
XX
```

```
CC   The present sequence encodes a maize cellulose synthase polypeptide.
CC   The cellulose synthase can be used for the improvement of stalk quality
CC   for improved stand or silage. It also provides an increased concentration
CC   of cellulose in the pericarp, hardening the kernel and improving its
CC   handling ability. The sequences are used to produce transgenic plants
CC   and seeds expressing the cellulose synthase. The polynucleotide is
CC   used for modulating, preferably increasing, the level of the synthase
CC   in a plant cell. The plants are preferably monocots. The polynucleotide
CC   is also used as a probe or primer in the detection quantitation or
CC   isolation of gene transcripts. The probes are useful in detecting
CC   deficiencies in the level of mRNA in screenings for desired transgenic
CC   plant, for detecting mutations in the gene, for monitoring upregulation
CC   of expression or changes in enzyme activity in screening assays of
CC   compounds, for detection of any number of allelic variants of the gene,
CC   or for use as molecular markers in plant breeding programs. The
CC   isolated nucleic acids of the present invention can also be used for
CC   recombinant expression of their encoded polypeptides or for use as
CC   immunogens in the preparation and/or screening of antibodies. The
CC   proteins can be employed in assays for enzyme agonists or antagonists
CC   of enzyme function or for use of immunogens or antigens to obtain
CC   antibodies specifically immunoreactive with a protein.
XX
SQ   Sequence 3704 BP; 909 A; 838 C; 962 G; 994 T; 1 other;

Alignment Scores:
Pred. No.:                0              Length:          3704
Score:                    3356.50        Matches:         637
Percent Similarity:       72.47%         Conservative:    145
Best Local Similarity:    59.04%         Mismatches:      170
Query Match:              64.06%         Indels:          127
DB:                       21             Gaps:            29

US-09-980-043-2 (1 978) x AAZ99533 (1-3704)
```





```
Qy      470 LysAlaValArgGluAlaMetCysIleLeuMetAspProGlnValGlyArgAspValCys 489
Db     1973 AAGGCTCTTCGAGAAGCTATGTGCTTCCTTATGGACCCAAACCTAGGAAGGAATGTCTGT 2032

Qy      490 TyrValGlnPheProGlnArgPheAspGlyIleAspLysSerAspArgTyrAlaAsnArg 509
Db     2033 TATGTCCAATTTCCTCAGAGGTTTGATGGTATTGATAGGAATGACCGATATGCAAACAGG 2092

Qy      510 AsnValValPhePheAspValAsnMetLysGlyLeuAspGlyIleGlnGlyProValTyr 529
Db     2093 AACACTGTGTTTTTCGATATTAACTTGAGAGGTCTTGACGGCATTCAAGGGCCAGTTTAT 2152

Qy      530 ValGlyThrGlyCysValPheAsnArgGlnAlaLeuTyrGlyTyrGlyProProSerMet 549
Db     2153 GTGGGAACTGGTTGTGTGTTTAACAGAACGGCCTTATATGGTTATGAGCCTCCA------ 2206

Qy      550 ProSerLeuArgLysArgLysAspSerSerSerCysPheSerCysCysCysProSerLys 569
Db     2207 ------GTCAAGAAAAAAAAG------CCAGGCTTCTTCTCTTCGCTTTGTGGGGGAAGG 2254

Qy      570 LysLysProAlaGlnAsp------ProAlaGluValTyrArgAspAlaLysArgGluAsp 587
Db     2255 AAAAAGACGTCAAAATCTAAGAAGAGCTCGGAAAAGAAGAAGTCACATAGACACGCAGAC 2314

Qy      588 LeuAsnAlaAlaIlePheAsnLeuThrGluIleAsp----------------AsnTyr 601
Db     2315 AGTTCTGTACCAGTATTTAATCTCGAAGATATAGAGGAAGGGATTGAAGGTTCTCAGTTT 2374

Qy      602 AspGluHisGluArgSerMetLeuIleSerGlnLeuSerPheGluLysThrPheGlyLeu 621
Db     2375 GATGAT---GAGAAATCGCTGATTATGTCTCAAATGAGCTTGGAGAAGAGATTTGGCCAG 2431

Qy      622 SerSerValPheIleGluSerThrLeuMetGluAsnGlyGlyValProGluSerAlaAsn 641
Db     2432 TCCAGTGTTTTTGTAGCCTCTACTCTGATGGAATATGGTGGTGTTCCACAATCTGCAACT 2491

Qy      642 SerProProPheIleLysGluAlaIleGlnValIleGlyCysGlyTyrGluGluLysThr 661
Db     2492 CCAGAGTCTCTTCTGAAAGAAGCTATTCATGTCATCAGCTGTGGCTATGAGGACAAAACT 2551

Qy      662 GluTrpGlyLysGlnIleGlyTrpIleTyrGlySerValThrGluAspIleLeuSerGly 681
Db     2552 GACTGGGGAACTGAGATTGGGTGGATCTATGGTTCTGTTACAGAAGACATTCTCACCGGA 2611

Qy      682 PheLysMetHisCysArgGlyTrpArgSerIleTyrCysMetProValArgProAlaPhe 701
Db     2612 TTCAAGATGCATGCTCGAGGCTGGCGATCAATCTACTGCATGCCTAAGCGACCAGCTTTC 2671

Qy      702 LysGlySerAlaProIleAsnLeuSerAspArgLeuHisGlnValLeuArgTrpAlaLeu 721
Db     2672 AAGGGATCTGCTCCTATCAACCTTTCGGATCGTTTGAATCAAGTGCTTCGGTGGGCTCTT 2731

Qy      722 GlySerValGluIlePhePheSerArgHisCysProLeuTrpTyrGlyPheGlyGlyGly 741
Db     2732 GGTTCCATTGAAATTCTTTTCAGCAGGCATTGTCCCATATGGTATGGCTAT---GGAGGC 2788

Qy      742 ArgLeuLysTrpLeuGlnArgLeuAlaTyrIleAsnThrIleValTyrProPheThrSer 761
Db     2789 CGGCTTAAATTCCTGGAGAGATTTGCTTATATCAACACAACAATTTATCCACTCACATCA 2848

Qy      762 LeuProLeuIleAlaTyrCysThrIleProAlaValCysLeuLeuThrGlyLysPheIle 781
Db     2849 ATCCCGCTCCTCCTGTACTGCATATTGCCAGCAGTTTGTCTTCTCACTGGGAAGTTCATC 2908

Qy      782 IleProThrLeuSerAsnLeuAlaSerMetLeuPheLeuGlyLeuPheIleSerIleIle 801
Db     2909 ATCCCAAAGATTAGTAACCTAGAGAGTGTTTGGTTTATATCGCTCTTTATCTCAATCTTT 2968

Qy      802 ValThrAlaValLeuGluLeuArgTrpSerGlyValSerIleGluAspLeuTrpArgAsn 821
Db     2969 GCCACTGGTATCCTTGAGATGAGGTGGAGTGGTGTTGGCATTGATGAATGGTGGAGGAAC 3028

Qy      822 GluGlnPheTrpValIleGlyGlyValSerAlaHisLeuPheAlaValPheGlnGlyPhe 841
```

```
Db     3029 GAGCAGTTCTGGGTCATTGGTGGTATTTCTGCGCATTTATTTGCCGTCTTCCAGGGTCTC 3088

Qy      842 LeuLysMetLeuAlaGlyIleAspThrAsnPheThrValThrAlaLysAlaAla---Glu 860
Db     3089 CTGAAGGTGCTTGCTGGTATCGACACGAGCTTCACTGTCACCTCTAAGGCCACTGACGAA 3148

Qy      861 AspAlaGluPheGlyGluLeuTyrMetValLysTrpThrThrLeuLeuIleProProThr 880
Db     3149 GAAGGTGATTTTGCTGAGCTCTACATGTTCAAGTGGACAACGCTTCTGATCCCACCAACC 3208

Qy      881 ThrLeuLeuIleIleAsnMetSerGly--CysAlaGlyPheSerAspAlaLeuAsnLys 899
Db     3209 ACTATTTTGATCATCAACCTGGTCGGCGTGGTCGCTGGCATTTCCTACGCAATCAATAGC 3268

Qy      900 GlyTyrGluAlaTrpGlyProLeuPheGlyLysValPhePheAlaPheTrpValIleLeu 919
Db     3269 GGTTACCAGTCATGTGGACCTCTTTTCGGGAAGTCTTCTTTTGCGTTCTGGGTGATTGTC 3328

Qy      920 HisLeuTyrProPheLeuLysGlyLeuMetGlyArgGlnAsnLeuThrProThrIleVal 939
Db     3329 CACCTCTATCCTTTCTTAAAGGGCTTATGGGAAGACAGAACCGCACGCCGACCATTGTC 3388

Qy      940 ValLeuTrpSerValLeuLeuAlaSerValPheSerLeuValTrpValLysIleAsnPro 959
Db     3389 GTTGTCTGGGCTATCTCCTTGCGTCGATCTTTTCCCTGATGTGGGTTCGTATCGATCCA 3448

Qy      960 PheValAsnLysValAspAsnThrLeuValAlaGluThrCysIleSerIleAspCys 978
Db     3449 TTCACCACCCGGGTCACTGGCCCTGATATCGCAAA    --TGTGGCATCAACTGC 3493
```

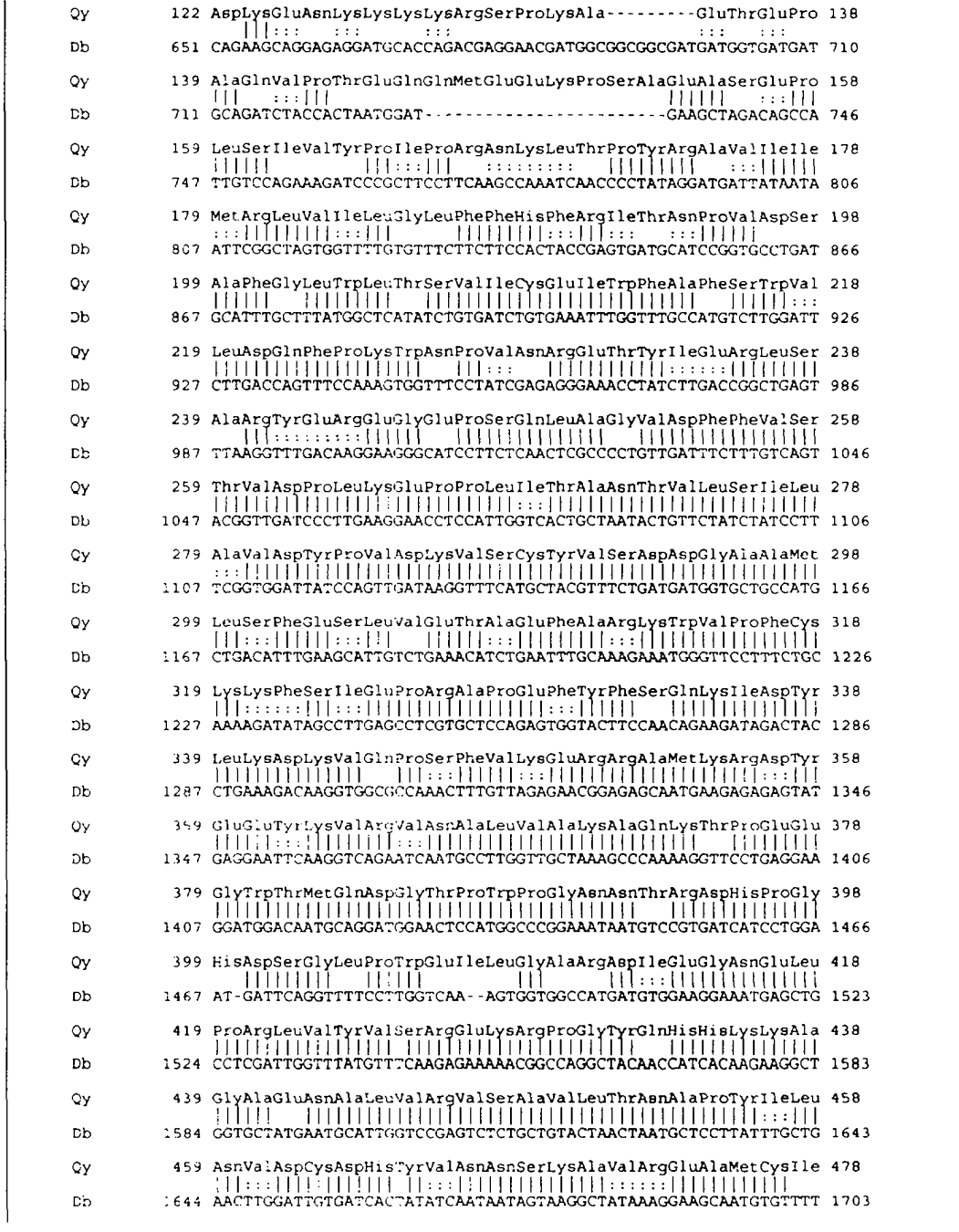
RESULT 14
AAZ99491
ID AAZ99491 standard; DNA; 3568 BP.
XX
AC AAZ99491;
XX
DT 03-JUL-2000 (first entry)
XX
DE DNA encoding a maize cellulose synthase.
XX
KW Maize; cellulose synthase; stalk quality; stand; silage; cellulose;
KW transgenic plant; plant breeding marker; ss.
XX
OS Zea mays.
XX
FH Key Location/Qualifiers
FT CDS 63..3236
FT /*tag= a
FT /product= "cellulose synthase"
FT /note= "no termination codon given"
XX
PN WO200009706-A2.
XX
PD 24-FEB 2000.
XX
PF 16-AUG 1999; 99WO-US18760.
XX
PR 17-AUG-1998; 98US-0096822.
XX
PA (PION-) PIONEER HI BRED INT INC.
XX
PI Dhugga KS, Helentjaris TG, Bowen BA, Wang X;
XX
DR WPI; 2000-224343/19.
DR P-PSDB; AAY84107.
XX
PT New genes which encode maize cellulose synthase polypeptides in plants
PT useful for modulating the expression of cellulose synthase in plants
PT and to produce transgenic plants expressing the novel protein -
XX
PS Claim 1; Page 82-87; 119pp; English.
XX

```
CC   The present sequence encodes a maize cellulose synthase polypeptide.
CC   The cellulose synthase can be used for the improvement of stalk quality
CC   for improved stand or silage. It also provides an increased concentration
CC   of cellulose in the pericarp, hardening the kernel and improving its
CC   handling ability. The sequences are used to produce transgenic plants
CC   and seeds expressing the cellulose synthase. The polynucleotide is
CC   used for modulating, preferably increasing, the level of the synthase
CC   in a plant cell. The plants are preferably monocots. The polynucleotide
CC   is also used as a probe or primer in the detection quantitation or
CC   isolation of gene transcripts. The probes are useful in detecting
CC   deficiencies in the level of mRNA in screenings for desired transgenic
CC   plant, for detecting mutations in the gene, for monitoring upregulation
CC   of expression or changes in enzyme activity in screening assays of
CC   compounds, for detection of any number of allelic variants of the gene,
CC   or for use as molecular markers in plant breeding programs. The
CC   isolated nucleic acids of the present invention can also be used for
CC   recombinant expression of their encoded polypeptides or for use as
CC   immunogens in the preparation and/or screening of antibodies. The
CC   proteins can be employed in assays for enzyme agonists or antagonists
CC   of enzyme function or for use of immunogens or antigens to obtain
CC   antibodies specifically immunoreactive with a protein.
XX
SQ   Sequence 3568 BP; 879 A; 825 C; 941 G; 922 T; 1 other;

Alignment Scores:
Pred. No.:                0                 Length:          3568
Score:                    3353.00           Matches:         630
Percent Similarity:       72.54%            Conservative:    162
Best Local Similarity:    58.44%            Mismatches:      177
Query Match:              63.99%            Indels:          121
DB:                       21                Gaps:            19

US-09-980-043-2 (1-978) x AAZ99491 (1-3568)
```





```
Qy      479 LeuMetAspProGlnValGlyArgAspValCysTyrValGlnPheProGlnArgPheAsp 498
Db     1704 ATGATGGATCCTTTGCTTGGAAAGAAAGTTTGCTATGTGCAGTTTCCTCAAAGATTTGAT 1763

Qy      499 GlyIleAspLysSerAspArgTyrAlaAsnArgAsnValValPhePheAspValAsnMet 518
Db     1764 GGGATTGATCGCCATGATCGATATGCTAACAGAAATGTTGTCTTTTTCGATATCAACATG 1823

Qy      519 LysGlyLeuAspGlyIleGlnGlyProValTyrValGlyThrGlyCysValPheAsnArg 538
Db     1824 AAAGGTTTGGATGGTATCCAGGGCCCAATTTATGTGGGTACTGGATGTGTCTTCAGAAGG 1883

Qy      539 GlnAlaLeuTyrGlyTyrGlyProProSerMetProSerLeuArgLysArgLysAspSer 558
Db     1884 CAGGCATTATATGGCTACGATGCTCCCAAA----------ACAAAGAAGCCACCATCA 1931

Qy      559 SerSer--------------------CysPheSerCysCysCys--- ------- -- 566
Db     1932 AGAACTTGCAACTGCTGGCCAAAGTGGTGCATTTGCTGTTGCTGTTTTGGTAACAGGAAG 1991

Qy      567 ProSerLysLysLysProAlaGlnAspPro------AlaGluValTyrArgAspAlaLys 584
Db     1992 ACCAAGAAGAAGACCAAGACCTCTAAACCTAAATTTGAGAAGATAAAGAAACTTTTTAAG 2051

Qy      585 ArgGluAspLeuAsnAlaAlaIlePheAsnLeuThrGluIleAspAsnTyr--------- 601
Db     2052 AAAAAGGAAAATCAAGCCCCTGCATATGCTCTTGGTGAAATTGATGAAGCCGCTCCAGGA 2111

Qy      602 AspGluHisGluArgSerMetLeuIleSerGlnLeuSerPheGluLysThrPheGlyLeu 621
Db     2112 GCTGAAAATGAAAAGGCTAGTATTGTAAATCAACAGAAGTTGGAAAAGAAATTTGGCCAG 2171

Qy      622 SerSerValPheIleGluSerThrLeuMetGluAsnGlyGlyValProGluSerAlaAsn 641
Db     2172 TCTTCAGTTTTTGTTGCATCCACACTTCTTGAGAATGGTGGAACCCTGAAGAGTGCCAGT 2231

Qy      642 SerProProPheIleLysGluAlaIleGlnValIleGlyCysGlyTyrGluGluLysThr 661
Db     2232 CCAGCTTCTCTTCTGAAGGAAGCTATACATGTCATCAGTTGTGGATATGAAGACAAAACA 2291

Qy      662 GluTrpGlyLysGlnIleGlyTrpIleTyrGlySerValThrGluAspIleLeuSerGly 681
Db     2292 GGCTGGGGAAAAGATATTGGTTGGATTTATGGATCAGTCACAGAAGATATTCTTACTGGG 2351

Qy      682 PheLysMetHisCysArgGlyTrpArgSerIleTyrCysMetProValArgProAlaPhe 701
Db     2352 TTTAAGATGCACTGCCATGGTTGGCGGTCAATTTACTGCATACCTAAACGGGCCGCCTTC 2411

Qy      702 LysGlySerAlaProIleAsnLeuSerAspArgLeuHisGlnValLeuArgTrpAlaLeu 721
Db     2412 AAAGGTTCCGCACCTCTCAATCTTTCCGATCGTTTTCACCAGGTTCTTCGGTGGGCTCTT 2471

Qy      722 GlySerValGluIlePhePheSerArgHisCysProLeuTrpTyrGlyPheGlyGlyGly 741
Db     2472 GGTTCAATTGAAATTTTGTTCAGCAACCACTGCCCTCTCTGGTATGGGTATGGTGGTGGA 2531

Qy      742 ArgLeuLysTrpLeuGlnArgLeuAlaTyrIleAsnThrIleValTyrProPheThrSer 761
Db     2532 ---CTAAAGTTCCTGGAAAGGTTTTCGTACATTAACTCCATCGTATACCCTTGGACATCT 2588

Qy      762 LeuProLeuIleAlaTyrCysThrIleProAlaValCysLeuLeuThrGlyLysPheIle 781
Db     2589 ATCCCGCTCTTGGCCTATTGCACATTGCCTGCCATCTGCTTGCTGACAGGGAAATTTATC 2648

Qy      782 IleProThrLeuSerAsnLeuAlaSerMetLeuPheLeuGlyLeuPheIleSerIleIle 801
Db     2649 ACGCCAGAGCTTAACAATGTTGCCAGCCTCTGGTTCATGTCACTTTTCATCTGCATTTTT 2708

Qy      802 ValThrAlaValLeuGluLeuArgTrpSerGlyValSerIleGluAspLeuTrpArgAsn 821
Db     2709 GCTACGAGCATCCTGGAAATGAGATGGAGTGGTGTAGGCATCGATGACTGGTGGAGAAAC 2768

Qy      822 GluGlnPheTrpValIleGlyGlyValSerAlaHisLeuPheAlaValPheGlnGlyPhe 841
Db     2769 GAGCAGTTTTGGGT ATTGGAGGGGTGTCTTCACATCTCTTTGCTGTGTTCCAGGGACTC 2828

Qy      842 LeuLysMetLeuAl GlyIleAspTh AsnPheThrValThrAlaLysAlaAlaGluAsp 861
Db     2829 CTCAAGGTCATAGC GGTGTAGACAC GACGCTTCACTGTGACATCCAAGGGCGGAGACGAC 2888

Qy      862 AlaGluPheGlyG. euTyrMetValLysTrpThrThrLeuLeuIleProProThrThr 881
Db     2889 GAGGAGTTCTCAGA TGTACACATTCAAATGGACGACCCTTCTGATACCTCCGACAACC 2948

Qy      882 LeuLeuIleIleAs MetSerGly- -CysAlaGlyPheSerAspAlaLeuAsnLysGly 900
Db     2949 CTGCTCCTACTGAA TCATTGGAGCGGTAGCTGGCATCTCCAATGCGATCAACAACGGA 3008

Qy      901 TyrGluAlaTrpG ProLeuPheGlyLysValPheIleAlaPheTrpValIleLeuHis 920
Db     3009 TATGAATCAGGTT TGTTTTTAAACT TTTTTTGCATTTTGGGTGATCGTCGCAT 3068

Qy      921 LeuCysPr ProPr yrAspMetArAr AsnLeuThrProThrIleValVal 940
Db     3069 CTTTACCC TCCT AAATTTCTTTTG AGAAGAACAGGACGCCAACGATTGTCATT 3128

Qy      941 LeuTrpSerValLe LeuAlaSerValPheSerLeuValTrpValLysIleAsnProPhe 960
Db     3129 GTCTGGTCCATCCT TGGCTTCGATCTTCTCTTGCTTTGGGTCCGGATCGACCCGTTC 3188

Qy      961 ValAsnLysValAs AsnThrLeuValAlaGlnThrCysIleSerIleAspCys 978
Db     3189 CTTGCGAAGGATGA GTTCCCTGTTGGAGGA------TGTGGTCTGGATTGC 3236

RESULT 15
AAC49550
ID   AAC49550 standard; DNA; 3198 BP.
XX
AC   AAC49550;
XX
DT   18-OCT-2000 (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 61572.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway;
KW   metabolic pathway; promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2
XX
PD   06-SEP-2000.
XX
PF   25 FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121925.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0125264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
```

PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY 1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19 MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY 1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03 JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18 JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN 1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN 1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN 1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30 JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01 JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055
PR 06 JUL-1999; 99US-0142390
PR 08 JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL 1999; 99US-0144085.
PR 16 JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19 JUL-1999; 99US-0144334.
PR 19 JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20 JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21 JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.

```
R   21-OCT-1999;   99US-0160815.
R   22-OCT-1999;   99US-0160980.
R   22-OCT-1999;   99US-0160981.
R   22-OCT-1999;   99US-0160989.
R   25-OCT-1999;   99US-0161404.
R   25-OCT-1999;   99US-0161405.
R   25-OCT-1999;   99US-0161406.
R   26-OCT-1999;   99US-0161359.
R   26-OCT-1999;   99US-0161360.
R   26-OCT-1999;   99US-0161361.
R   28-OCT-1999;   99US-0161920.
R   28-OCT-1999;   99US-0161992.
R   28-OCT-1999;   99US-0161993.
R   29-OCT-1999;   99US-0162142.

lignment Scores:
red. No.:              0              Length:        3198
core:                  3350.00        Matches:       654
ercent Similarity:     71.94%         Conservative:  123
est Local Similarity:  60.56%         Mismatches:    185
uery Match:            63.93%         Indels:        118
B:                     21             Gaps:          21

S-09-980-043-2 (1-978) x AAC49550 (1-3198)

y          2 MetGluSer--------------GlyAlaProIle--------------------Cys 9
             ||||||||               |||   |||:::                     :||
b          1 ATGGAATCCGAAGGAGAAACCGCGGGAAAGCCGATGAAGAACATTGTTCCGCAGACTTGC 60

y         10 HisThrCysGlyGluGlnValGlyHisAspAlaAsnGlyGluLeuPheValAlaCysHis 29
               |||   :::   ||||||        :::|||:::   ||||||||||||
b         61 CAGATCTGTAGTGACAATGTTGGCAAGACTGTTGATGGAGATCGTTTTGTGGCTTGTGAT 120

y         30 GluCysSerTyrProMetCysLysSerCysPheGluPheGluIleAsnGluGlyArgLys 49
               ||||||:::|||:::|||:::   |||:::|||:::|||      :::|||   :::
b        121 ATTTGTTCATTCCCAGTTTGTCGGCCTTGCTACGAGTATGAGAGGAAAGATGGGAATCAA 180

y         50 ValCysLeuArgCys----------------------GlySerPro------------ 57
               |||   :::|||                        ||||||||
b        181 TCTTGTCCTCAGTGCAAAACCAGATACAAGAGGCTCAAAGGTAGTCCTGCTATTCCTGGT 240

y         58 ---TyrAspGluAsnLeuLeuAspAspVal------------------------------ 66
                  ||||||:::   |||   |||
b        241 GATAAAGACGAGGATGGCTTAGCTGATGAAGGTACTGTTGAGTTCAACTACCCTCAGAAG 300

y         67 ------------------------------------------------GluLysLys 69
                                                             |||:::
b        301 GAGAAAATTTCAGAGCGGATGCTTGGTTGGCATCTTACTCGTGGGAAGGGAGAGGAAATG 360

y         70 GlySerGlyAsnGlnSerThrMetAlaSerHis----------LeuAsnAspSer 84
               |||                    ||||||              |||
b        361 GGGGAACCCCAGTATGATAAAGAGGTCTCTCACAATCATCTTCCTCGTCTCACGAGCAGA 420

y         85 GlnAspVal--------------GlyIleHisAlaArgHisIleSerSerValSerThr 99
               ||||||                        :::|:|      ||||||
b        421 CAAGATACTTCAGGAGAGTTTTCTGCTGCCTCACCTGAACGCCTCTCTGTATCTTCTACT 480

y        100 Val----------------------AspSerGluMetAsnAspGlu--- -------- 107
               :::                    |||:::::|||
b        481 ATCGCTGGGGGAAAGCGCCTTCCCTATTCATCAGATGTCAATCAATCACCAAATAGAAGG 540

y        108 -----------------TyrGlyAsnProIleTrpLysAsnArgValLysSerCysLys 121
                              ||||||     ||||||   ||||||          |||
b        541 ATTGTGGATCCTGTTGGACTCGGGAATGTAGCTTGGAAGGAGAGAGTTGATGGCTGGAAA 600

y        122 AspLysGluAsnLysLysLys---LysArgSerProLysAlaGluThrGluProAlaGln 140
               |||:::   |||              |||   :::|||   :::|||
b        601 ATGAAGCAAGAGAAGAATACTGGTCCTGTCAGCACGCAGGCTGCTTCTGAAAGAGGTGGA 660

y        141 ValProThrGluGlnGlnMet--------------GluGluLysProSerAlaGluAla 155
               |||      :::                    |||         :::   ||||||
b        661 GTAGATATTGATGCCAGCACAGATATCCTAGCAGATGAGGCTCTGCTGAATGACGAAGCG 720
```

```
Qy       156 SerGluProLeuSerIleValTyrProIleProArgAsnLysLeuThrProTyrArgAla 175
             :::|| :||||:             ||.'||   :::::::::   |||||||.|
Db       721 AGGCAGCGTCTGTCAAGGAAAGTTTCAATTCCTTCATCACGGATCAATCCTTACAGAATG 780

Qy       176 ValIleIleMetArgLeuValIleLeuGlyLeuPheHisPheArgIleThrAsnPro 195
             |||:|:::::|:|||||||||||.|   :||||:   |||:::|||:|||||||||.
Db       781 GTTATTATGCTGCGGCTTGTTATCCTTGTCTCTTCTTGCATTACCGTATAACAAACCCA 840

Qy       196 ValAspSerAlaPheGlyLeuTrpLeuThrSerValIleCysGluIleTrpPheAlaPhe 215
             |||   :::|'|||    | :|||||   :|||||||||||||||||||||||||||
Db       841 GTGCCAAATGCCTTTGCTCTATGGCTGGTCTCTGTGATATGTGAGATCTGGTTTGCCTTA 900

Qy       216 SerTrpValLeuAspGlnPheProLysTrpAsnProValAsnArgGluThrTyrIleGlu 235
             ||||'|::: :||| :|||||. |||' |||   ||||:|||||.||||||||:::::
Db       901 TCCTGGATTTTGGATCAGTTTCCCAAGTGGTTTCCTGTGAACCGTGAAACCTACCTCGAC 960

Qy       236 ArgLeuSerAlaAs:TyrGluAs:PheGlyGluProSerGlnLeuAlaGlyValAspPhe 255
             :.|||'::: :|  |  :::  ||. |.|  ||||:|||||:|||  :.||||
Db       961 AGGCTTCTTTAAATCTACTAA:AAATGAA:TGA:::ATCACAGTTAGCTGCTGTTGACATT 1020

Qy       256 PheValSerThrVa.AspProGlnLys:luProProLeuIleThrAlaAsnThrValLeu 275
             || |||| |||  |||||'||:: ||| .'||| ::|||:::||||:|:|||':.|||||
Db      1021 TTCGTGAGTACTGTTGACCCCTTGAAGGAGCCACCCCTTGTGACAGCCAACACAGTGCTC 1080

Qy       276 SerIleLeuAlaValAspTyrProValAspLysValSerCysTyrValSerAspAspGly 295
             |||:|||'|||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1081 TCTATTCTGGCTGTTGACTACCCAGTTGACAAGGTGTCCTGTTATGTTTCTGATGATGGT 1140

Qy       296 AlaAlaMetLeuSerPheGluSerLeuValGlnThrAlaGluPheAlaArgLysTrpVal 315
             |||||||||||||| ||||:||.|:|  |:||||:::|||||||||'|||||:||
Db      1141 GCTGCTATGTTATCATTTGAATCACTTGCAGAAACATCAGAGTTTGCTCGTAAATGGGTA 1200

Qy       316 ProPheCysLysLysPheSerIleGluProArgAlaProGluPheTyrPheSerGlnLys 335
             |||||||||||||| :::|||||||:|||'|||||:|||||:::||||||:::   |||
Db      1201 CCATTTTGCAAGAAATATAGCATAGAGCCTCGTGCACCAGAATGGTACTTTGCTGCGAAA 1260

Qy       336 IleAspTyrLeuLysAspLysValGlnProSerPheValLysGluArgArgAlaMetLys 355
             |||||||||||||||.||||'||||||  ||.|||||||||:::|||||||||||||||
Db      1261 ATAGATTACTTGAAGGATAAAGTTCAGACATCATTTGTCAAAGATCGTAGAGCTATGAAG 1320

Qy       356 ArgAspTyrGluGluTyrLysValArgValAsnAlaLeuValAlaLysAlaGlnLysThr 375
             |||:::|||:|||||':::|||:::|||:::|:|||||||||:::|||||   |||
Db      1321 AGGGAATATGAGGAATTTAAAATCCGAATCAATGCACTTGTTTCCAAAGCCCTAAAATGT 1380

Qy       376 ProGluGluGlyTrpThrMetGlnAspGlyThrProTrpProGlyAsnAsnThrArgAsp 395
             ||||::|||:|||||  |||||||.'|||||'|||||:|||||:|||||:|||||||||
Db      1381 CCTGAAGAAGGGTGGGTTATGCAAGATGGCACACCGTGGCCTGGAAATAATACAAGGGAC 1440

Qy       396 HisPro--------------------  GlyHisAspSerGlyLeuProTrpGluIleLeu 408
             ||||||                     :||   :::   ||||||
Db      1441 CATCCAGGAATGATCCAGGTGTTCTTAGGGCAAAATGGTGGACTT------ -------- 1485

Qy       409 GlyAlaArgAspIleGluGlyAsnGluLeuProArgLeuValTyrValSerArgGluLys 428
                 ||:   : |||. ||| !|||| ||||||||||||||||:|||||||||||
Db      1486 ---    --GATGCAGAGGGCAATGAGCTCCCGCGTTTGGTATATGTTTCTCGAGAAAAG 1536

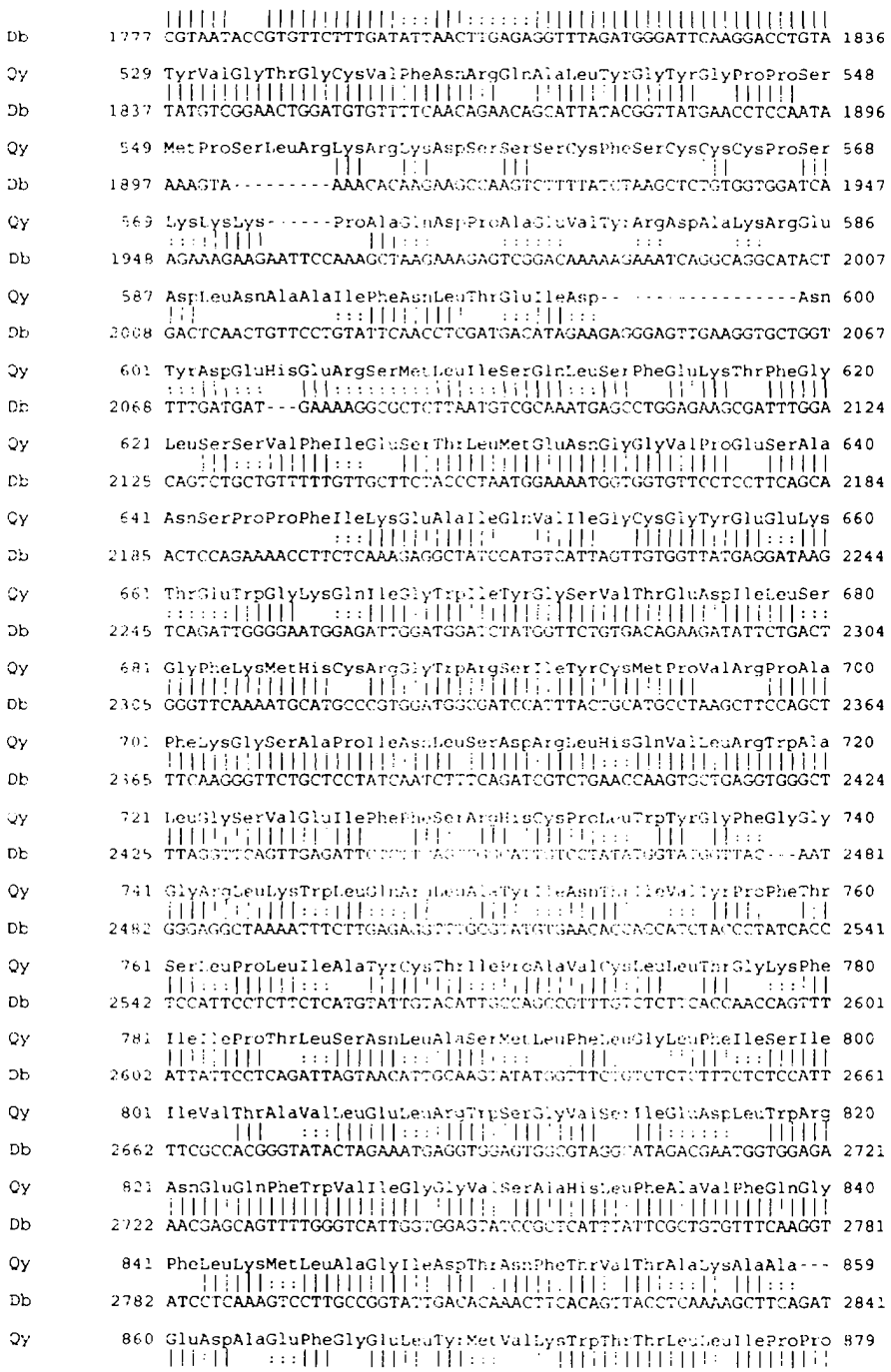
Qy       429 ArgProGlyTyrGlnHisHisLysLysAlaGlyAlaGluAsnAlaLeuValArgValSer 448
             |||||.'||:::|||'||||:|||:|||||' ||   |||||||||||||||||||||
Db      1537 CGACCAGGATTCCAACACCACAAAAAGGCTGGTGCTATGAATGCACTGGTGAGAGTTTCA 1596

Qy       449 AlaValLeuThrAsnAlaProTyrIleLeuAsnValAspCysAspHisTyrValAsnAsn 468
             ||||||.||||'|||   |||:::|||:|:::|||||||||||||||:::||||||||
Db      1597 GCAGTTCTTACCAATGGACCTTTCATCTTGAATCTTGATTGTGATCATTACATAAATAAC 1656

Qy       469 SerLysAlaValArgGluAlaMetCysIleLeuMetAspProGlnValGlyArgAspVal 488
             ||||||||:::|||||||||:::|||   |||||||||||   :::||::::   |||
Db      1657 AGCAAAGCCTTAAGAGAAGCAATGTGCTTCCTGATGGACCCAAACCTCGGGAAGCAAGTT 1716

Qy       489 CysTyrValGlnPheProGlnArgPheAspGlyIleAspLysSerAspArgTyrAlaAsn 508
             |||||||||.||||:|||||||||||||||||||||||||:::|||||||||||||||
Db      1717 TGTTATGTTCAGTTCCCACAAAGATTTGATGGTATCGATAAGAACGATAGATATGCTAAT 1776

Qy       509 ArgAsnValValPhePheAspValAsnMetLysGlyLeuAspGlyIleGlnGlyProVal 528
```



Search completed: October 31, 2003, 16:00:29
Job time : 829.785 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame plus_p2n model

Run on: October 31, 2003, 15:18:06 ; Search time 10526.3 Seconds
(without alignments)
9800.930 Million cell updates/sec

Title: US-09-980-043-2
Perfect score: 5240
Sequence: 1 MMESGAPICHTCGEQVGHPA.... ....PFVNKVLNTLVAEMQISIDC 978

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2898711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5797422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09980043/runat_29102003_163852_21863/app_query.fasta_1.1358
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOCAL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09980043_@CGN_1_1_6944_@runat_29102003_163852_21863 -NCPU=6 -ICPU=3
NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGJOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	5240	100.0	3232	8	AF072131	AF072131 Populus t
2	4540	86.6	3186	8	GHU58283	U58283 Gossypium h
3	4540	86.6	3328	6	AR164434	AR164434 Sequence
4	4329.5	82.6	2958	8	AF267742	AF267742 Arabidops
5	4322.5	82.5	3109	8	AY099636	AY099636 Arabidops
6	4220	80.5	3145	8	AF323039	AF323039 Zinnia el
7	3990	76.1	6886	8	AF413210	AF413210 Gossypium
c 8	3926	74.9	94301	8	ATF28A21	AL035526 Arabidops
c 9	3926	74.9	199075	8	ATCHRIV49	AL161549 Arabidops
10	3465.5	66.1	3181	8	BT006111	BT006111 Arabidops
11	3465.5	66.1	3377	8	BT005710	BT005710 Arabidops
12	3464.5	66.1	3168	8	AF458083	AF458083 Arabidops
13	3401.5	64.9	162776	8	AP003237	AP003237 Oryza sat
14	3396	64.8	3395	8	AF081534	AF081534 Populus x
15	3366.5	64.2	3795	8	AF200533	AF200533 Zea mays
16	3366.5	64.2	3799	6	AX338680	AX338680 Sequence
17	3366	64.2	3723	8	AF150630	AF150630 Gossypium
18	3365.5	64.2	3745	8	AF200528	AF200528 Zea mays
19	3358.5	64.1	3222	6	AX652952	AX652952 Sequence
20	3356.5	64.1	3676	8	AF200529	AF200529 Zea mays
21	3353	64.0	3538	8	AF200530	AF200530 Zea mays
22	3350	63.9	3229	8	BT002335	BT002335 Arabidops
23	3350	63.9	3603	6	AR267557	AR267557 Sequence
24	3350	63.9	3603	6	AX030942	AX030942 Sequence
25	3350	63.9	3603	6	BD022676	BD022676 Manipulat
26	3350	63.9	3763	8	BT008654	BT008654 Arabidops
27	3347	63.9	3355	8	AY139754	AY139754 Arabidops
28	3346	63.9	3081	8	BT004543	BT004543 Arabidops
29	3346	63.9	3673	6	AR267560	AR267560 Sequence
30	3346	63.9	3673	6	AX030948	AX030948 Sequence
31	3346	63.9	3673	6	BD022679	BD022679 Manipulat
32	3345.5	63.8	3968	8	AF200531	AF200531 Zea mays
33	3345	63.8	3812	8	AF200532	AF200532 Zea mays
34	3343.5	63.8	3264	6	AX653232	AX653232 Sequence
35	3340	63.7	3081	8	AF088917	AF088917 Arabidops
36	3335.5	63.7	3752	8	AF200525	AF200525 Zea mays
37	3332	63.6	3725	8	AF200526	AF200526 Zea mays
38	3329	63.5	3614	6	AR267559	AR267559 Sequence
39	3329	63.5	3614	6	AX030946	AX030946 Sequence
40	3329	63.5	3614	6	BD022678	BD022678 Manipulat
41	3327	63.5	3682	8	AF027174	AF027174 Arabidops
42	3304	63.1	3277	8	AY095297	AY095297 Populus t
43	3277	62.5	3847	8	AF304374	AF304374 Nicotiana
44	3274	62.5	3255	6	AX505864	AX505864 Sequence
45	3274	62.5	3771	8	AF027173	AF027173 Arabidops

ALIGNMENTS

RESULT 1

```
F072131
OCUS        AF072131                3232 bp    mRNA    linear   PLN 26-APR-2001
EFINITION   Populus tremuloides secondary xylem cellulose synthase (celA) mRNA,
            complete cds.
CCESSION    AF072131
ERSION      AF072131.1  GI:4115904
EYWORDS     .
OURCE       Populus tremuloides (quaking aspen)
 ORGANISM   Populus tremuloides
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Malpighiales; Salicaceae; Populus.
EFERENCE    1  (bases 1 to 3232)
 AUTHORS    Wu,L., Joshi,C.P. and Chiang,V.L.
 TITLE      A xylem-specific cellulose synthase gene from aspen (Populus
            tremuloides) is responsive to mechanical stress
 JOURNAL    Plant J. 22 (6), 495-502 (2000)
 MEDLINE    20344632
  PUBMED    10886769
EFERENCE    2  (bases 1 to 3232)
 AUTHORS    Wu,L., Joshi,C.P. and Chiang,V.L.
 TITLE      Direct Submission
 JOURNAL    Submitted (14-JUN-1998) IWR/Forestry, Michigan Tech University,
            1400, Townsend Drive, Houghton, MI 49931, USA
EFERENCE    3  (bases 1 to 3232)
 AUTHORS    Wu,L., Joshi,C.P. and Chiang,V.L.
 TITLE      Direct Submission
 JOURNAL    Submitted (07-JAN-1999) IWR/Forestry, Michigan Tech University,
            1400, Townsend Drive, Houghton, MI 49931, USA
 REMARK     Sequence update by submitter
OMMENT      On Jan 8, 1999 this sequence version replaced gi:3243277.
EATURES             Location/Qualifiers
     source         1. .3232
                    /organism="Populus tremuloides"
                    /mol_type="mRNA"
                    /db_xref="taxon:3693"
                    /tissue_type="secondary xylem"
     gene           1. .3232
                    /gene="celA"
     CDS            69. .3005
                    /gene="celA"
                    /codon_start=1
                    /product="secondary xylem cellulose synthase"
                    /protein_id="AAD03417.1"
                    /db_xref="GI:4115905"
                    /translation="MMESGAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEI
                    NEGRKVCLRCGSPYDENLLDDVEKKGSGNQSTMASHLNDSQDVGIHARHISSVSTVDS
                    EMNDEYGNPIWKNRVKSCKDKENKKKKRSPKAETEPAQVPTEQQMEEKPSAEASEPLS
                    IVYPIPRNKLTPYRAVIIMRLVILGLFFHFRITNPVDSAFGLWLTSVICEIWFAFSWV
                    LDQFPKWNPVNRETYIERLSARYEREGEPSQLAGVDFFVSTVDPLKEPPLITANTVLS
                    ILAVDYPVDKVSCYVSDDGAAMLSFESLVETAEFARKWVPFCKKFSIEPRAPEFYFSQ
                    KIDYLKDKVQPSFVKERRAMKRDYEEYKVRVNALVAKAQKTPEEGWTMQDGTPWPGNN
                    TRDHPGHDSGLPWEILGARDIEGNELPRLVYVSREKRPGYQHHKKAGAENALVRVSAV
                    LTNAPYILNVDCDHYVNNSKAVREAMCILMDPQVGRDVCYVQFPQRFDGIDKSDRYAN
                    RNVVFFDVNMKGLDGIQGPVYVGTGCVFNRQALYGYGPPSMPSLRKRKDSSSCFSCCC
                    PSKKKPAQDPAEVYRDAKREDLNAAIFNLTEIDNYDEHERSMLISQLSFEKTFGLSSV
                    FIESTLMENGGVPESANSPPFIKEAIQVIGCGYEEKTEWGKQIGWIYGSVTEDILSGF
                    KMHCRGWRSIYCMPVRPAFKGSAPINLSDRLHQVLRWALGSVEIFFSRHCPLWYGFGG
                    GRLKWLQRLAYINTIVYPFTSLPLIAYCTIPAVCLLTGKFIIPTLSNLASMLFLGLFI
                    SIIVTAVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQGFLKMLAGIDTNFTVTA
                    KAAEDAEFGELYMVKWTTLLIPPTTLLIINMSGCAGFSDALNKGYEAWGPLFGKVFFA
                    FWVILHLYPFLKGLMGRQNLTPTIVVLWSVLLASVFSLVWVKINPFVNKVDNTLVAET
                    CISIDC"
ASE COUNT      895 a    668 c    772 g    897 t
RIGIN

lignment Scores:
red. No.:              0              Length:        3232
core:                  5240.00        Matches:       978
ercent Similarity:     100.00%        Conservative:  0
est Local Similarity:  100.00%        Mismatches:    0
uery Match:            100.00%        Indels:        0
B:                     8              Gaps:          0
```

```
US-09-980-043-2 (1-978) x AF072131 (1-3232)

Qy          1 MetMetGluSerGlyAlaProIleCysHisThrCysGlyGluGlnValGlyHisAspAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         69 ATGATGGAATCTGGAGCTCCTATATGCCATACCTGTGGTGAACAGGTGGGGCATGATGCA 128

Qy         21 AsnGlyGluLeuPheValAlaCysHisGluCysSerTyrProMetCysLysSerCysPhe 40
              ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        129 AATGGGGAGCTATTTGTGGCTTGCCATGAGTGTAGCTATCCCATGTGCAAGTCTTGTTTC 188

Qy         41 GluPheGluIleAsnGluGlyArgLysValCysLeuArgCysGlySerProTyrAspGlu 60
              ||||||||||||||  ||| |||||| |||||||||||||||||||||||||||||||||
Db        189 GAGTTTGAAATCAACGAGGGTCGGAAAGTTTGCTTGCGGTGTGGCTCGCCATATGATGAG 248

Qy         61 AsnLeuLeuAspAsnValGluLysLysGlySerGlyAsnGlnSerThrMetAlaSerHis 80
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        249 AACTTGCTTGATAACGTAGAAAAGAAGGGTTCTGGCAATCAATCCACAATGGCATCTCAC 308

Qy         81 LeuAsnAspSerIleAspValGlyLeuProLeuProThrIleSerSerValSerThrVal 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        309 CTCAACGATTCTATAGATGTTGGAATTCCATTGCCTACAATCAGCAGTGTTTCCACTGTG 368

Qy        101 AspSerGluMetAsnAspGluTyrGlyAsnProIleTrpLysAsnArgValLysSerCys 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        369 GATAGTGAAATGAACGATGAATATGGGAATCCAATTTGGAAGAATCGGGTGAAGAGCTGT 428

Qy        121 LysAspLysGluAsnLysLysLysLysArgSerProLysAlaGluThrGluProAlaGln 140
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        429 AAGGATAAAGAGAACAAGAAGAAAAAGAGAAGTCCTAAGGCTGAAACTGAACCAGCTCAA 488

Qy        141 ValProThrGluGlnGlnMetGluGluLysProSerAlaGluAlaSerGluProLeuSer 160
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        489 GTTCCTACAGAACAGCAGATGGAAGAGAAACCGTCTGCAGAGGCTTCGGAGCCGCTTTCA 548

Qy        161 IleValTyrProIleProArgAsnLysLeuThrProTyrArgAlaValIleIleMetArg 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        549 ATTGTTTATCCAATTCCACGCAACAAGCTCACACCATACAGAGCAGTGATCATTATGCGA 608

Qy        181 LeuValIleLeuGlyLeuPhePheHisPheArgIleThrAsnProValAspSerAlaPhe 200
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        609 CTGGTCATTCTGGGACTCTTCTTCCACTTCAGAATAACAAATCCTGTCGATAGTGCCTTT 668

Qy        201 GlyLeuTrpLeuThrSerValIleCysGluIleTrpPheAlaPheSerTrpValLeuAsp 220
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        669 GGCCTGTGGCTTACTTCTGTCATATGTGAGATCTGGTTTGCATTTTCTTGGGTGTTGGAT 728

Qy        221 GlnPheProLysTrpAsnProValAsnArgGluThrTyrIleGluArgLeuSerAlaArg 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        729 CAGTTCCCCAAGTGGAATCCTGTCAATAGAGAAACGTATATCGAAAGGCTGTCGGCAAGG 788

Qy        241 TyrGluArgGluGlyGluProSerGlnLeuAlaGlyValAspPhePheValSerThrVal 260
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        789 TATGAAAGAGAGGGTGAGCCTTCTCAGCTTGCTGGTGTGGATTTTTTCGTGAGTACTGTT 848

Qy        261 AspProLeuLysGluProProLeuIleThrAlaAsnThrValLeuSerIleLeuAlaVal 280
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        849 GATCCGCTGAAGGAACCGCCATTGATCACTGCCAATACAGTCCTTTCCATCCTTGCTGTG 908

Qy        281 AspTyrProValAspLysValSerCysTyrValSerAspAspGlyAlaAlaMetLeuSer 300
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        909 GACTATCCCGTCGATAAAGTCTCCTGCTACGTGTCTGATGATGGTGCAGCTATGCTTTCA 968

Qy        301 PheGluSerLeuValGluThrAlaGluPheAlaArgLysTrpValProPheCysLysLys 320
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        969 TTTGAATCTCTTGTAGAAACAGCTGAGTTTGCAAGGAAGTGGGTTCCGTTCTGCAAAAAA 1028

Qy        321 PheSerIleGluProArgAlaProGluPheTyrPheSerGlnLysIleAspTyrLeuLys 340
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1029 TTCTCAATTGAACCAAGAGCACCGGAGTTTTACTTCTCACAGAAAATTGATTACTTGAAA 1088

Qy        341 AspLysValGlnProSerPheValLysGluArgArgAlaMetLysArgAspTyrGluGlu 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1089 GACAAGGTTCAACCTTCTTTCGTGAAAGAACGTAGAGCAATGAAAAGGGATTATGAAGAG 1148

Qy        361 TyrLysValArgValAsnAlaLeuValAlaLysAlaGlnLysThrProGluGluGlyTrp 380

Db       1149 TACAAAGTCCGAGTTAATGCCCTGGTAGCAAAGGCTCAGAAAACACCTGAAGAAGGATGG 1208

Qy        381 ThrMetGlnAspGlyThrProTrpProGlyAsnAsnThrArgAspHisProGlyHisAsp 400

Db       1209 ACTATGCAAGATGGAACACCTTGGCCTGGGAATAACACACGGGATCACCCTGGGCATGAT 1268

Qy        401 SerGlyLeuProTrpGluIleLeuGlyAlaArgAspIleGlyGlyAsnGluLeuProArg 420

Db       1269 TCAGGTCTTCCTTGGGAAATACTGGGAGCTCGTGACATTGAGGGAAATGAACTACCTCGT 1328

Qy        421 LeuValTyrValSerArgGluLysArgProGlyTyrGlnHisHisLysLysAlaGlyAla 440

Db       1329 CTAGTATATGTCTCCAGGGAGAAGAGACCTGGCTACCAGCACACAAAAAGGCTGGTGCA 1388

Qy        441 GluAsnAlaLeuValArgValSerAlaValLeuThrAsnAlaProTyrIleLeuAsnVal 460

Db       1389 GAAAATGCTCTGGTGAGAGTGTCTGCAGTACTCACAAATGCTCCCTACATCCTCAATGTT 1448

Qy        461 AspCysAspHisTyrValAsnAsnSerLysAlaValArgGluAlaMetCysIleLeuMet 480

Db       1449 GATTGTGATCACTATGTAAACAATAGCAAGGCTGTTCGAGAAGCAATGTGCATCCTGATG 1508

Qy        481 AspProGlnValGlyArgAspValCysTyrValGlnPhePheGlnArgPheAspGlyIle 500

Db       1509 GACCCACAAGTAGGTCGAGATGTATGCTATGTGCAGTTCCCTCAGAGGTTTGATGGCATA 1568

Qy        501 AspLysSerAspArgTyrAlaAsnArgAsnValValPhePheAspValAsnMetLysGly 520

Db       1569 GATAAGAGTGATCGCTACGCCAATCGTAACGTAGTTTTCTTTGATGTTAACATGAAAGGG 1628

Qy        521 LeuAspGlyIleGlnGlyProValTyrValGlyThrGlyCysValPheAsnArgGlnAla 540

Db       1629 TTGGATGGCATTCAAGGACCAGTATACGTAGGAACTGGTTGTGTTTTCAACAGGCAAGCA 1688

Qy        541 LeuTyrGlyTyrGlyProProSerMetProSerLeuArgLysArgLysAspSerSerSer 560

Db       1689 CTTTACGGCTACGGGCCTCCTTCTATGCCCAGCTTACGCAAGAGAAAGGATTCTTCATCC 1748

Qy        561 CysPheSerCysCysCysProSerLysLysLysProAlaGluAspProAlaGluValTyr 580

Db       1749 TGCTTCTCATGTTGCTGCCCCTCAAAGAAGAAGCCTGCTCAAGATCCAGCTGAGGTATAC 1808

Qy        581 ArgAspAlaLysArgGluAspLeuAsnAlaAlaIlePheAspLeuThrGluIleAspAsn 600

Db       1809 AGAGATGCAAAAAGAGAGGATCTCAATGTTGCATATTCAA TTACAGAGATTGATAAT 1868

Qy        601 TyrAspGluHisGluArgSerMetLeuIleSerGlnLeuSerProGluLysThrPheGly 620

Db       1869 TATGACGAGCATGAAAGGTCAATGCTGATCTCCCAGTTGAGCTTGAGAAAACTTTTGGC 1928

Qy        621 LeuSerSerValPheIleGluSerThrLeuMetGluAsnGlyGlyValProGluSerAla 640

Db       1929 TTATCTTCTGTCTTCATTGAGTCTACACTAATGGAGAATGGAGGAGTACCCGAGTCTGCC 1988

Qy        641 AsnSerProProPheIleLysGluAlaIleGlnValIleGlyCysGlyTyrGluGluLys 660

Db       1989 AACTCACCACCATTCATCAAGGAAGCGATTCAAGTCATCGGTTGTGGCTATGAAGAGAAG 2048

Qy        661 ThrGluTrpGlyLysGlnIleGlyTrpIleTyrGlySerValThrGluAspIleLeuSer 680

Db       2049 ACTGAATGGGGAAAACAGATTGGTTGGATATATGGGTCAGTGACTGAGGATATCTTAAGT 2108

Qy        681 GlyPheLysMetHisCysArgGlyTrpArgSerIleTyrCysMetProValArgProAla 700

Db       2109 GGCTTCAAGATGCACTGCCGAGGATGGAGATCAATTTACTGCATGCCCGTAAGGCCTGCA 2168

Qy        701 PheLysGlySerAlaProIleAsnLeuSerAspArgLeuHisGlnValLeuArgTrpAla 720

Db       2169 TTCAAAGGATCTGCACCCATCAACCTGTCTGATAGATTGCACCAGGTCCTCCGATGGGCT 2228

Qy        721 LeuGlySerValGluIlePhePheSerArgHisCysProLeuTrpTyrGlyPheGlyGly 740

Db       2229 CTTGGTTCTGTGGAAATTTTCTTTAGCAGACACTGTCCCCTCTGGTACGGGTTTGGAGGA 2288

Qy        741 GlyArgLeuLysTrpLeuGlnArgLeuAlaTyrIleAsnThrIleValTyrProPheThr 760

Db       2289 GGCCGTCTTAAATGGCTCCAAAGGCTTGCGTATATAAACACCATTGTGTACCCATTTACA 2348

Qy        761 SerLeuProLeuIleAlaTyrCysThrIleProAlaValCysLeuLeuThrGlyLysPhe 780

Db       2349 TCCCTCCCTCTCATTGCCTATTGCACAATTCCTGCAGTTTGTCTGCTCACCGGAAAATTC 2408

Qy        781 IleIleProThrLeuSerAsnLeuAlaSerMetLeuPheLeuGlyLeuPheIleSerIle 800

Db       2409 ATCATACCAACGCTCTCAAACCTGGCAAGCATGCTGTTTCTTGGCCTCTTTATCTCCATC 2468

Qy        801 IleValThrAlaValLeuGluLeuArgTrpSerGlyValSerIleGluAspLeuTrpArg 820

Db       2469 ATTGTAACTGCGGTGCTTGAGCTAAGATGGAGCGGTGTCAGCATTGAAGATTTATGGCGT 2528

Qy        821 AsnGluGlnPheTrpValIleGlyGlyValSerAlaHisLeuPheAlaValPheGlnGly 840

Db       2529 AATGAACAATTCTGGGTGATCGGAGGTGTTTCAGCCCATCTCTTTGCGGTCTTCCAGGGA 2588

Qy        841 PheLeuLysMetLeuAlaGlyIleAspThrAsnPheThrValThrAlaLysAlaAlaGlu 860

Db       2589 TTCTTAAAAATGTTGGCTGGCATCGATACGAACTTCACTGTCACAGCAAAAGCAGCCGAA 2648

Qy        861 AspAlaGluPheGlyGluLeuTyrMetValLysTrpThrThrLeuLeuIleProProThr 880

Db       2649 GATGCAGAATTTGGGGAGCTATATATGGTCAAGTGGACAACACTTTTGATTCCTCCAACC 2708

Qy        881 ThrLeuLeuIleIleAsnMetSerGlyCysAlaGlyPheSerAspAlaLeuAsnLysGly 900

Db       2709 ACACTTCTCATTATCAATATGTCGGGTTGTGCTGGATTCTCTGATGCACTCAACAAAGGA 2768

Qy        901 TyrGluAlaTrpGlyProLeuPheGlyLysValPhePheAlaPheTrpValIleLeuHis 920

Db       2769 TATGAAGCATGGGGGCCTCTCTTTGGCAAGGTGTTCTTTGCTTTCTGGGTGATTCTTCAT 2828

Qy        921 LeuTyrProPheLeuLysGlyLeuMetGlyArgGlnAsnLeuThrProThrIleValVal 940

Db       2829 CTCTATCCATTCCTTAAAGGTCTAATGGGTCGCCAAAACCTAACACCAACCATTGTTGTT 2888

Qy        941 LeuTrpSerValLeuLeuAlaSerValPheSerLeuValTrpValLysIleAsnProPhe 960

Db       2889 CTCTGGTCAGTGCTGTTGGCCTCTGTCTTCTCTCTCGTTTGGGTCAAGATCAATCCATTC 2948

Qy        961 ValAsnLysValAspAsnThrLeuValAlaGluThrCysIleSerIleAspCys 978

Db       2949 GTTAACAAAGTTGATAACACCTTGGTTGCGGAGACCTGCATTTCCATTGATTGC 3002
```

```
RESULT 2
GHU58283
LOCUS       GHU58283                3186 bp    mRNA    linear   PLN 05-DEC-1996
DEFINITION  Gossypium hirsutum cellulose synthase (celA1) mRNA, complete cds.
ACCESSION   U58283
VERSION     U58283.1  GI:1706955
KEYWORDS    .
SOURCE      Gossypium hirsutum (upland cotton)
  ORGANISM  Gossypium hirsutum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium.
REFERENCE   1  (bases 1 to 3186)
  AUTHORS   Pear,J.R., Kawagoe,Y., Schreckengost,W.E., Delmer,D.P. and
            Stalker,D.M.
  TITLE     Higher plants contain homologs of the bacterial celA genes encoding
            the catalytic subunit of cellulose synthase
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 93 (22), 12637-12642 (1996)
  MEDLINE   97057296
   PUBMED   8901635
```

```
REFERENCE   2  (bases 1 to 3186)
  AUTHORS   Pear,J.R., Kawagoe,Y., Schreckengost,W.E., Delmer,D.P. and
            Stalker,D.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-MAY-1996) J.R. Pear, Calgene, Inc, 1920 5th Street,
            Davis, CA 95616, USA
FEATURES             Location/Qualifiers
     source          1. .3186
                     /organism="Gossypium hirsutum"
                     /mol_type="mRNA"
                     /strain="Acala SJ-2"
                     /db_xref="taxon:3635"
                     /tissue_type="fiber"
                     /dev_stage="17-21 days post-anthesis"
     gene            1. .3186
                     /gene="celA1"
     CDS             77. .3001
                     /gene="celA1"
                     /codon_start=1
                     /product="cellulose synthase"
                     /protein_id="AAB37766.1"
                     /db_xref="GI:1706956"
                     /translation="MMESGVPVCHTCGEHVGLNVNGEPFVACHECNFPICKSCFEYDL
                     KEGRKACLRCGSPYDENLLDDVEKATGDQSTMAAHLNKSQDVGIHARHISSVSTLDSE
                     MAEDNGNSIWKNRVESWKEKKNKKKKPATTKVEREAEIPPEQQMEDKPAPDASQPLST
                     IIPIPKSRLAPYRTVIIMRLIILGLFFHYRVTNPVDSAFGLWLTSVICEIWFAFSWVL
                     DQFPKWYPVNRETYIDRLSARYEREGEPDELAAVDFFVSTVDPLKEPPLITANTVLSI
                     LALDYPVDKVSCYISDDGAAMLTFESLVETADFARKWVPFCKKFSIEPRAPEFYFSQK
                     IDYLKDKVQPSFVKERRAMKRDYEEYKIRINALVAKAQKTPDEGWTMQDGTSWPGNNP
                     RDHPGMIQVFLGYSGARDIEGNELPRLVYVSREKRPGYQHHKKAGAENALVRVSAVLT
                     NAPFILNLDCDHYVNNSKAVREAMCFLMDPQVGRDVCYVQFPQRFDGIDRSDRYANRN
                     TVFFDVNMKGLDGIQGPVYVGTGCVFNRQALYGYGPPSMPSFPKSSSSSCSCCCPGKK
                     EPKDPSELYRDAKREELDAAIFNLREIDNYDEYERSMLISQTSFEKTFGLSSVFIEST
                     LMENGGVAESANPSTLIKEAIHVISCGYEEKTAWGKEIGWIYGSVTEDILTGFKMHCR
                     GWRSIYCMPLRPAFKGSAPINLSDRLHQVLRWALGSVEIFLSRHCPLWYGFGGGRLKW
                     LQRLAYINTIVYPFTSLPLIAYCSLPAICLLTGKFIIPTLSNLASVLFLGLFLSIIVT
                     AVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQGFLKMLAGIDTNFTVTAKAADD
                     ADFGELYIVKWTTLLIPPTTLLIVNMVGVVAGFSDALNKGYEAWGPLFGKVFFSFWVI
                     LHLYPFLKGLMGRQNRTPTIVVLWSVLLASVFSLVWVRINPFVSTADSTTVSQSCISI
                     DC"
BASE COUNT      864 a    654 c    723 g    945 t
ORIGIN

Alignment Scores:
Seq. No.:               0          Length:        3186
Score:                  4540.00    Matches:       845
Percent Similarity:     93.27%     Conservative:  69
Best Local Similarity:  86.22%     Mismatches:    59
Query Match:            86.64%     Indels:        8
DB:                     8          Gaps:          6

US-09-980-043-2 (1-978) x GHU58283 (1-3186)

Qy          1 MetMetGluSerGlyAlaProIleCysHisThrCysGlyGluGlnValGlyHisAspAla 20
              ||||||||||||||||   |||:::|||||||||||||||||   ||||||   :::
Db         77 ATGATGGAATCTGGGGTTCCTGTTTGCCACACTTGTGGTGAACATGTTGGGTTGAATGTT 136

Qy         21 AsnGlyGluLeuPheValAlaCysHisGluCysSerTyrProMetCysLysSerCysPhe 40
              |||||||||   ||||||||||||||||||||||:::::|||:::||||||||||||||||
Db        137 AATGGTGAACCTTTTGTGGCTTGCCATGAATGTAATTTCCCTATTTGTAAGAGTTGTTTT 196

Qy         41 GluPheGluIleAsnGluGlyArgLysValCysLeuArgCysGlySerProTyrAspGlu 60
              |||:::::::::   ||||||||||||   ||||||||||||||||||||||||||||||
Db        197 GAGTATGATCTTAAGGAAGGACGAAAAGCTTGCTTGCGTTGTGGTAGTCCATATGATGAA 256

Qy         61 AsnLeuLeuAspAspValGluLysLysGlySerGlyAsnGlnSerThrMetAlaSerHis 80
              ||||||||||||||||||   |||   :::|||:::||||||||||||||||||:::|||
Db        257 AACCTGTTGGACGATGTCGAG---AAGGCCACCGGCGATCAATCGACAATGGCTGCACAT 313

Qy         81 LeuAsnAspSerGlnAspValGlyIleHisAlaArgHisIleSerSerValSerThrVal 100
              ||||||   ||||||||||||||||||||||||||||||||||||||||||||||:::
Db        314 TTGAACAAGTCTCAGGATGTTGGAATTCATGCAAGACATATCAGCAGTGTGTCTACATTG 373

Qy        101 AspSerGluMetAs AspGluTyrGlyAsnPro IleTrpLysAsnArgValLysSer Cys 120
              |||||||.|||.   :::::   |||. |   ; ||||||||||||||||:::|||
Db        374 GATAGTGAAATGGCTGAAGACAATGGGAATTCGATTTGGAAGAACAGGGTGGAAAGTTGG 433

Qy        121 LysAspLysGluAs LysLysLysLysArgSerProLysAlaGluThrGluProAlaGln 140
              |||:::|||:::|| ||||||:|||.   :::                 |||:::
Db        434 AAAGAAAAGAAGAA AAGAAGAAGAAGCCTGCAACAACTAAGGTTGAAAGAGAGGCTGAA 493

Qy        141 ValProThrGluGl GlnMetGluGluLysProSerAlaGluAlaSerGluProLeuSer 160
              :::|.   |||| ||||||:|||:|||| :::   :::||||||:::|||||||
Db        494 ATCCCACCTGAGCA ACAAATGGAAGATAAACCTGGCACCGGATGCTTCCCAGCCCCTCTCG 553

Qy        161 IleValTyrProI ProArgAsnLysLeuTh ProTyrArgAlaValIleIleMetArg 180
              :::   ||||. |||:::::::::||.   |||||||||   |||:|||||||||
Db        554 ACTATAATTCCAA CCGAAAAGCAGACTTGCACCATACCGAACCGTGATCATTATGCGA 613

Qy        181 LeuValIleLeuG LeuPhePheHisPheArgIleThrAsnProValAspSerAlaPhe 200
              |||:::  |||  | |||  ||  :::|||||||||||||||||||||||||||||
Db        614 TTGATCATTCTTG TTTTTCTTCCATTTCAGAATAACAAACCCCGTTGACAGTGCTTTT 673

Qy        201 GlyLeuTrpLeuT SerValIleCysHisIleTrpPheAlaPheSerTrpValLeuAsp 220
              |||||.||||.|  |||  |||.|||  ||||||||||||||||||||||||||||||
Db        674 GGACTGTGGCTCA TTCAGTCATATGTGAAATCTGGTTTGCATTTTCCTGGGTGTTGGAT 733

Qy        221 GlnPheProLysTrpAsnProValAsnArgGluThrTyrIleGluArgLeuSerAlaArg 240
              ||||||||||||   ||||||||||||||||||||||||||||||||||||||||||||
Db        734 CAGTTCCCTAAGTG GTATCCTGTTAACAGGGAAACATACATTGACAGACTATCTGCAAGA 793

Qy        241 TyrGluArgGluGlyGluProSerGlnLeuAlaGlyValAspPhePheValSerThrVal 260
              ||||||||||||||  ||||:   :::||||:   |||||||||||||||||||||||||
Db        794 TATGAAAGAGAAGG GAACCTGATGAACTTGCTGCAGTTGACTTCTTCGTGAGTACAGTG 853

Qy        261 AspProLeuLysGl ProProLeuIleThrAlaAsnThrValLeuSerIleLeuAlaVal 280
              |||||||||||||  |||:|||||||.|||:|||||.|||||||||||||||:||||:::
Db        854 GATCCATTGAAAGA CCTCCATTGATTACTGCCAATACTGTGCTTTCCATCCTTGCCTTG 913

Qy        281 AspTyrProValAsnLysValSerCysTyrValSerAspAspGlyAlaAlaMetLeuSer 300
              |||:||||||||:|||||||||||||||||||||||:|||  ||||||||||||||:::
Db        914 GACTACCCGGTGGA AAGGTCTCTTGTTATATATCTGATGATGGTGCGGCCATGCTGACA 973

Qy        301 PheGluSerLeuValGluThrAlaGluPheAlaArgLysTrpValProPheCysLysLys 320
              |||||||.|||:|||||||||||||||||||||||||||||||||||||||||||||||
Db        974 TTTGAATCTCTAGTAGAAACAGCCGACTTTGCAAGAAAGTGGGTTCCATTCTGCAAAAAA 1033

Qy        321 PheSerIleGluP pArgAlaProGluPheTyrPheSerGlnLysIleAspTyrLeuLys 340
              |||||:|||||||: |||| ||||.|||||||||||||||||||||||||||||||||||
Db       1034 TTTTCCATTGAACC CGGGCACCTGAGTTTTACTTCTCACAGAAGATTGATTACTTGAAA 1093

Qy        341 AspLysValGlnPr SerPheValLysGluArgArgAlaMetLysArgAspTyrGluGlu 360
              |||||:|||:|||| ||||||||||||||.|||||||||||||||||||||||||||||
Db       1094 GATAAAGTGCAGCC TCTTTTGTAAAAGAACGTAGAGCTATGAAAAGAGATTATGAAGAG 1153

Qy        361 TyrLysValArgVa AsnAlaLeuValAlaLysAlaGlnLysThrProGluGluGlyTrp 380
              ||||||:::|||:: ||||||||||:|||:|||||||||||||||||||:::||||||||
Db       1154 TACAAAATTCGAAT AATGCTTTAGTTGCAAAGGCTCAGAAAACACCTGATGAAGGATGG 1213

Qy        381 ThrMetGlnAspGlyThrProTrpProGlyAsnAsnThrArgAspHisProGlyHisAsp 400
              |||||||||||||:|||:   |||||||||||||||||||||||||||||||||||||||
Db       1214 ACAATGCAAGATGGAACTTCTTGGCCAGGAAATAACCCGCGTGATCACCCTGG-CATGAT 1272

Qy        401 SerGlyLeuProTrpGluIle LeuGlyAlaArgAspIleGluGlyAsnGluLeuProAr 420
              |||||   ||||:   |||    ||||||:|||||||||||||||||||||||||:||
Db       1273 TCAGGTTTTCCTTGG-- ATATAGTGGTGCTCGTGACATCGAAGGAAATGAACTTCCTCG 1329

Qy        420 gLeuValTyrValSerArgGluLysArgProGlyTyrGlnHisHisLysLysAlaGlyAl 440
              ||||:||||:||||.|||||||||||||||||||||||||||||||||||||||||:||
Db       1330 ACTGGTTTACGTCTCTAGAGAGAAGAGACCTGGCTACCAACACCACAAAAAGGCTGGTGC 1389

Qy        440 aGluAsnAlaLeuValArgValSerAlaValLeuThrAsnAlaProTyrIleLeuAsnVa 460
              |||||||||||.|||:|||||||||||.|||||||||||||||||||||||:::||:||
Db       1390 TGAAAATGCTTTGGTTAGGGTGTCTGCAGTTCTTACAAATGCTCCCTTCATCCTCAATCT 1449

Qy        460 lAspCysAspHisTyrValAsnAsnSerLysAlaValArgGluAlaMetCysIleLeuMe 480
```

```
Db       1450 TGATTGTGACCACTATGTTAACAATAGCAAGGCAGTTAGGGAGGCAATGTGCTTCTTGAT 1509

Qy        480 tAspProGlnValGlyArgAspValCysTyrValGlnPheProGlnArgPheAspGlyIl 500

Db       1510 GGACCCACAAGTTGGTCGAGATGTATGCTATGTGCAGTTTCCTCAAAGATTTGATGGCAT 1569

Qy        500 eAspLysSerAspArgTyrAlaAsnArgAsnValValPheIleAspValAsnMetLysGl 520

Db       1570 AGATAGGAGTGATCGATATGCCAATAGGAACACAGTTTTCTTTGATGTTAACATGAAAGG 1629

Qy        520 yLeuAspGlyIleGlnGlyProValTyrValGlyThrGlyCysValPheAsnArgGlnAl 540

Db       1630 TCTTGATGGAATCCAAGGGCCAGTTTATGTGGGAACAGGTTTGTTTTCAATAGGCAAGC 1689

Qy        540 aLeuTyrGlyTyrGlyProProSerMetProSerLeuArgLysArgLysAspSerSerSe 560

Db       1690 ACTTTATGGCTATGGTCCACCTTCAATGCCAAGTTTTCCCAAG---TCATCCTCCTCATC 1746

Qy        560 rCysPheSerCysCysCysProSerLysLysLysProAlaGlnAspProAlaGluValTy 580

Db       1747 TTGC---TCGTGTTGCTGCCCGGGCAAGAAGGAACCT---AAAGATCCATCAGAGCTTTA 1800

Qy        580 rArgAspAlaLysArgGluAspLeuAsnAlaAlaIlePheAsnLeuThrGluIleAspAs 600

Db       1801 TAGGGATGCAAAACGGGAAGAACTTGATGCTGCCATCTTTAACCTTAGGGAAATTGACAA 1860

Qy        600 nTyrAspGluHisGluArgSerMetLeuIleSerGlnLeuSerPheGluLysThrPheGl 620

Db       1861 TTATGATGAGTATGAAAGATCAATGTTGATCTCTCAAACAAGCTTTGAGAAAACTTTTGG 1920

Qy        620 yLeuSerSerValPheIleGluSerThrLeuMetGluAsnGlyGlyValProGluSerAl 640

Db       1921 CTTATCTTCAGTCTTCATTGAATCTACACTAATGGAGAATGAGGAGTGGCTGAATCTGC 1980

Qy        640 aAsnSerProProPheIleLysGluAlaIleGlnValIleGlyCysGlyTyrGluGluLy 660

Db       1981 CAACCCTTCCACACTAATCAAGGAAGCAATTCATGTCATCACTGTGGCTATGAAGAGAA 2040

Qy        660 sThrGluTrpGlyLysGlnIleGlyTrpIleTyrGlySerValThrGluAspIleLeuSe 680

Db       2041 GACTGCATGGGGGAAAGAGATTGGATGGATATATGGTTCAGTCACTGAGGATATCTTAAC 2100

Qy        680 rGlyPheLysMetHisCysArgGlyTrpArgSerIleTyrCysMetProValArgProAl 700

Db       2101 CGGCTTCAAAATGCACTGCCGAGGATGGAGATCGATTTACTATGCCTTAAGGCCAGC 2160

Qy        700 aPheLysGlySerAlaProIleAsnLeuSerAspArgLeuGlnGlnValLeuArgTrpAl 720

Db       2161 ATTGAAAGGATCTGCACCCATGAATCTCCGAATTATTAGATCTTCGATGGGC 2220

Qy        720 aLeuGlySerValGluIlePheLeuSerArgHisCysProLeuTrpTyrGlyPheGlyGl 740

Db       2221 TCTTGGATCTGTTGAAATTTTCCTAAGCAGGCATTGCCCTCATGGTATGGCTTTGGAGG 2280

Qy        740 yGlyArgLeuLysTrpLeuGlnArgLeuAlaTyrIleAsnLeuValGlyTyrProPheTh 760

Db       2281 TGGTCGTCTTAAATGGCTTCAAAGACTAGCATATATAAAGACATTGTCTATCCTTTCAC 2340

Qy        760 rSerLeuProLeuIleAlaTyrCysThrIleProAlaValCysLeuLeuThrGlyLysPh 780

Db       2341 ATCCCTTCCACTCATTGCCTATTGTTCACTACCAGCAATCTTCGTTCTCACAGGAAAATT 2400

Qy        780 eIleIleProThrLeuSerAsnLeuAlaSerMetLeuPheLeuGlyLeuPheIleSerIl 800

Db       2401 TATCATACCAACGCTCTCAAACCTGGCAAGTGTTCTCTTTCGGCCTTTTCCTTTCCAT 2460

Qy        800 eIleValThrAlaValLeuGluLeuArgTrpSerGlyValSerIleGluAspLeuTrpAr 820

Db       2461 TATCGTGACTGCTGTTCTCGAGCTCCGATGGAGTGGTGTCATATTGAGGACTTATGGCG 2520

Qy        820 gAsnGluGlnPheTrpValIleGlyGlyValSerAlaHisLeuPheAlaValPheGlnGl 840

Db       2521 TAACGAGCAGTTTTGGGTCATCGGTGGCGTTTCAGCCCATCTCTTTGCCGTCTTCCAAGG 2580

Qy        840 yPheLeuLysMetLeuAlaGlyIleAspThrAsnPheThrValThrAlaLysAlaAlaGl 860

Db       2581 TTTCCTTAAGATGCTTGCGGGCATTGACACCAACTTTACTGTCACTGCCAAAGCAGCTGA 2640

Qy        860 uAspAlaGluPheGlyGluLeuTyrMetValLysTrpThrThrLeuLeuIleProProTh 880

Db       2641 TGATGCAGATTTTGGTGAGCTCTACATTGTGAAATGGACTACACTTCTAATCCCTCCAAC 2700

Qy        880 rThrLeuLeuIleIleAsnMetSerGly---CysAlaGlyPheSerAspAlaLeuAsnLy 899

Db       2701 AACACTCCTCATCGTCAACATGGTTGGTGTCGTTGCCGGATTCTCCGATGCCCTCAACAA 2760

Qy        899 sGlyTyrGluAlaTrpGlyProLeuPheGlyLysValPhePheAlaPheTrpValIleLe 919

Db       2761 AGGGTACGAAGCTTGGGGACCACTCTTTGGCAAAGTGTTCTTTTCCTTCTGGGTCATCCT 2820

Qy        919 uHisLeuTyrProPheLeuLysGlyLeuMetGlyArgGlnAsnLeuThrProThrIleVa 939

Db       2821 CCATCTTTATCCATTCCTCAAAGGTCTTATGGGACGCCAAAACAGGACACCAACCATTGT 2880

Qy        939 lValLeuTrpSerValLeuLeuAlaSerValPheSerLeuValTrpValLysIleAsnPr 959

Db       2881 TGTCCTTTGGTCAGTGTTGTTGGCTTCTGTCTTCTCTCTTGTTTGGGTTCGGATCAACCC 2940

Qy        959 oPheValAsnLysValAspAsnThrLeuValAlaGluThrCysIleSerIleAspCys 978

Db       2941 GTTTGTCAGCACCGCCGATAGCACCACCGTGTCACAGAGCTGCATTTCCATTGATTGT 2998
```

```
RESULT 3
AR164434
LOCUS       AR164434                3328 bp    DNA     linear   PAT 17-OCT-2001
DEFINITION  Sequence 1 from patent US 6271443.
ACCESSION   AR164434
VERSION     AR164434.1  GI:16235654
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 3328)
  AUTHORS   Stalker,D.M., Pear,J.R. and Delmer,D.
  TITLE     Cotton and rice cellulose synthase DNA sequences
  JOURNAL   Patent: US 6271443-A 1 07-AUG-2001;
FEATURES             Location/Qualifiers
     source          1. .3328
                     /organism="unknown"
BASE COUNT      914 a    689 c    763 g    962 t
ORIGIN

Alignment Scores:
Pred. No.:                0              Length:          3328
Score:                    4540.00        Matches:         845
Percent Similarity:       93.27%         Conservative:    69
Best Local Similarity:    86.22%         Mismatches:      59
Query Match:              86.64%         Indels:          8
DB:                       6              Gaps:            6

US-09-980-043-2 (1 978) x AR164434 (1-3328)

Qy          1 MetMetGluSerGlyAlaProIleCysHisThrCysGlyGluGlnValGlyHisAspAla 20

Db        179 ATGATGGAATCTGGGGTTCCTGTTTGCCACACTTGTGGTGAACATGTTGGGTTGAATGTT 238

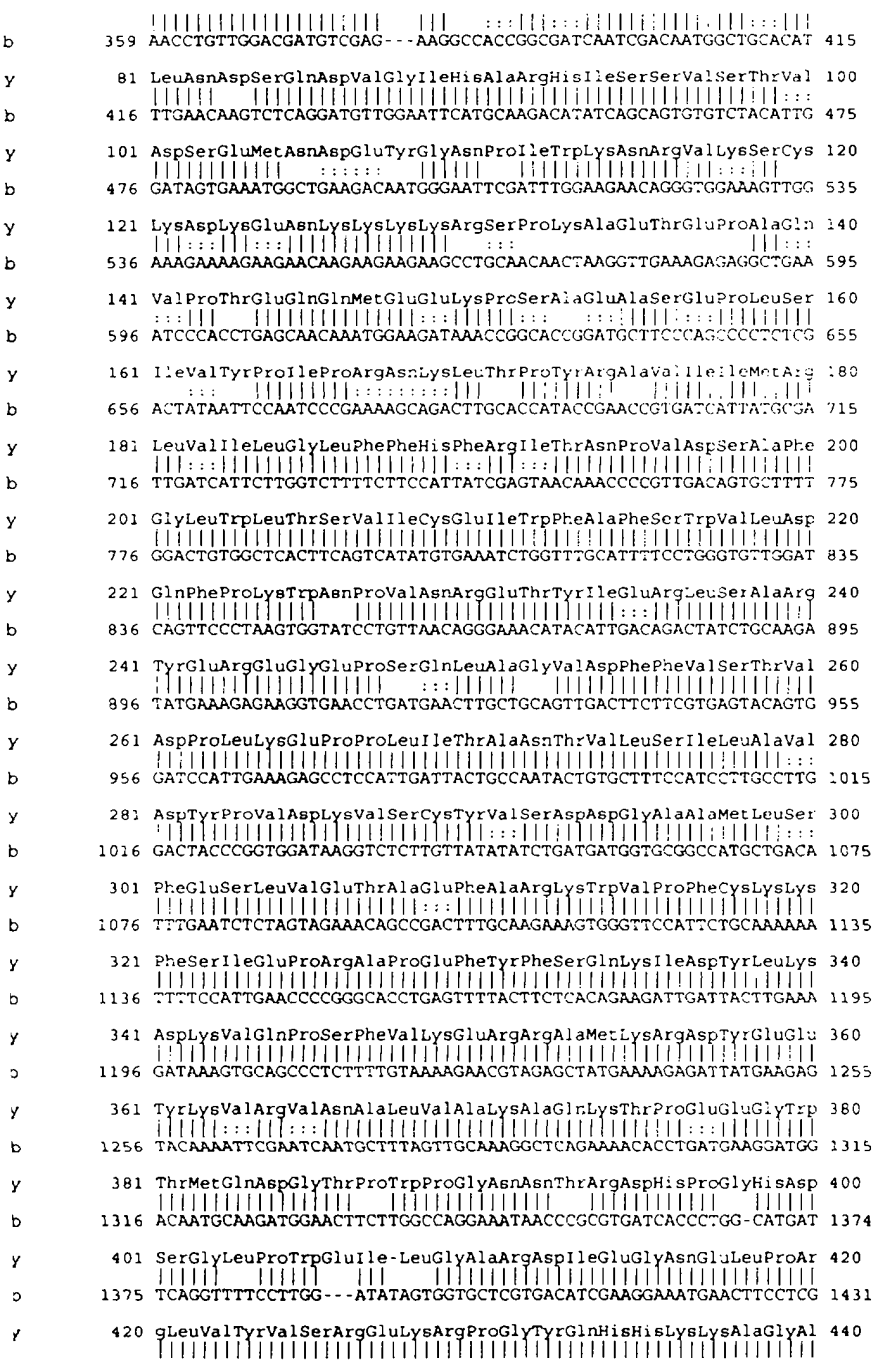
Qy         21 AsnGlyGluLeuPheValAlaCysHisGluCysSerTyrProMetCysLysSerCysPhe 40

Db        239 AATGGTGAACCTTTTGTGGCTTGCCATGAATGTAATTTCCCTATTTGTAAGAGTTGTTTT 298

Qy         41 GluPheGluIleAsnGluGlyArgLysValCysLeuArgCysGlySerProTyrAspGlu 60

Db        299 GAGTATGATCTTAAGGAAGGACGAAAAGCTTGCTTGCGTTGTGGTAGTCCATATGATGAA 358

Qy         61 AsnLeuLeuAspAspValGluLysLysGlySerGlyAsnGlnSerThrMetAlaSerHis 80
```


b 359 AACCTGTTGGACGATGTCGAG---AAGGCCACCGGCGATCAATCGACAATGGCTGCACAT 415
y 81 LeuAsnAspSerGlnAspValGlyIleHisAlaArgHisIleSerSerValSerThrVal 100
b 416 TTGAACAAGTCTCAGGATGTTGGAATTCATGCAAGACATATCAGCAGTGTGTCTACATTG 475
y 101 AspSerGluMetAsnAspGluTyrGlyAsnProIleTrpLysAsnArgValLysSerCys 120
b 476 GATAGTGAAATGGCTGAAGACAATGGGAATTCGATTTGGAAGAACAGGGTGGAAAGTTGG 535
y 121 LysAspLysGluAsnLysLysLysLysArgSerProLysAlaGluThrGluProAlaGln 140
b 536 AAAGAAAAGAAGAACAAGAAGAAGAAGCCTGCAACAACTAAGGTTGAAAGAGAGGCTGAA 595
y 141 ValProThrGluGlnGlnMetGluGluLysProSerAlaGluAlaSerGluProLeuSer 160
b 596 ATCCCACCTGAGCAACAAATGGAAGATAAACCGGCACCGGATGCTTCCCAGCCCCTCTCG 655
y 161 IleValTyrProIleProArgAsnLysLeuThrProTyrArgAlaValIleIleMetArg 180
b 656 ACTATAATTCCAATCCCGAAAAGCAGACTTGCACCATACCGAACCGTGATCATTATGCGA 715
y 181 LeuValIleLeuGlyLeuPhePheHisPheArgIleThrAsnProValAspSerAlaPhe 200
b 716 TTGATCATTCTTGGTCTTTTCTTCCATTATCGAGTAACAAACCCCGTTGACAGTGCTTTT 775
y 201 GlyLeuTrpLeuThrSerValIleCysGluIleTrpPheAlaPheSerTrpValLeuAsp 220
b 776 GGACTGTGGCTCACTTCAGTCATATGTGAAATCTGGTTTGCATTTTCCTGGGTGTTGGAT 835
y 221 GlnPheProLysTrpAsnProValAsnArgGluThrTyrIleGluArgLeuSerAlaArg 240
b 836 CAGTTCCCTAAGTGGTATCCTGTTAACAGGGAAACATACATTGACAGACTATCTGCAAGA 895
y 241 TyrGluArgGluGlyGluProSerGlnLeuAlaGlyValAspPhePheValSerThrVal 260
b 896 TATGAAAGAGAAGGTGAACCTGATGAACTTGCTGCAGTTGACTTCTTCGTGAGTACAGTG 955
y 261 AspProLeuLysGluProProLeuIleThrAlaAsnThrValLeuSerIleLeuAlaVal 280
b 956 GATCCATTGAAAGAGCCTCCATTGATTACTGCCAATACTGTGCTTTCCATCCTTGCCTTG 1015
y 281 AspTyrProValAspLysValSerCysTyrValSerAspAspGlyAlaAlaMetLeuSer 300
b 1016 GACTACCCGGTGGATAAGGTCTCTTGTTATATATCTGATGATGGTGCGGCCATGCTGACA 1075
y 301 PheGluSerLeuValGluThrAlaGluPheAlaArgLysTrpValProPheCysLysLys 320
b 1076 TTTGAATCTCTAGTAGAAACAGCCGACTTTGCAAGAAAGTGGGTTCCATTCTGCAAAAAA 1135
y 321 PheSerIleGluProArgAlaProGluPheTyrPheSerGlnLysIleAspTyrLeuLys 340
b 1136 TTTTCCATTGAACCCCGGGCACCTGAGTTTTACTTCTCACAGAAGATTGATTACTTGAAA 1195
y 341 AspLysValGlnProSerPheValLysGluArgArgAlaMetLysArgAspTyrGluGlu 360
b 1196 GATAAAGTGCAGCCCTCTTTTGTAAAAGAACGTAGAGCTATGAAAAGAGATTATGAAGAG 1255
y 361 TyrLysValArgValAsnAlaLeuValAlaLysAlaGlnLysThrProGluGluGlyTrp 380
b 1256 TACAAAATTCGAATCAATGCTTTAGTTGCAAAGGCTCAGAAAACACCTGATGAAGGATGG 1315
y 381 ThrMetGlnAspGlyThrProTrpProGlyAsnAsnThrArgAspHisProGlyHisAsp 400
b 1316 ACAATGCAAGATGGAACTTCTTGGCCAGGAAATAACCCGCGTGATCACCCTGG-CATGAT 1374
y 401 SerGlyLeuProTrpGluIle-LeuGlyAlaArgAspIleGluGlyAsnGluLeuProAr 420
b 1375 TCAGGTTTTCCTTGG---ATATAGTGGTGCTCGTGACATCGAAGGAAATGAACTTCCTCG 1431
y 420 gLeuValTyrValSerArgGluLysArgProGlyTyrGlnHisHisLysLysAlaGlyAl 440

Db 1492 ACTGGTTTACGTC[illegible]AAGAAAAAAAGCTGCTACCAACACCACAAAAAGGCTGGTGC 1491
Qy 440 aGluAsnAlaLeu[illegible]ArgValSerAlaValLeuThrAsnAlaProTyrIleLeuAsnVa 460
Db 1492 TGAAAATGCTTTG[illegible]AGGGTGTCTGCAGTTCTTACAAATGCTCCCTTCATCCTCAATCT 1551
Qy 460 lAspCysAspHisT[illegible]yrValAsnAsnSerLysAlaValArgGluAlaMetCysIleLeuMe 480
Db 1552 TGATTGTGACCACT[illegible]GTTAACAATAGCAAGGCAGTTAGGGAGGCAATGTGCTTCTTGAT 1611
Qy 480 rAspProGlnVal[illegible]yArgAspValCysTyrValGlnPheProGlnArgPheAspGlyIl 500
Db 1612 GGACCCACAAGTT[illegible]TCGAGATGTATGCTATGTGCAGTTTCCTCAAAGATTTGATGGCAT 1671
Qy 500 eAspLysSer[illegible]AspA[illegible]gTyrAlaAsnA[illegible]rgAsnValValPhePheAspValAsnMetLysGl 520
Db 1672 AGATAG[illegible]AGTGAT[illegible]ATATGCCAATA[illegible]GAACACAGTTTTCTTTGATGTTAACATGAAAGG 1731
Qy 520 yLeuAs[illegible]yL[illegible]yPr[illegible]yTyr[illegible]yAsnGlyCysValPheAsnArgGlnAl 540
Db 1732 TCTTGA[illegible]AAC[illegible]AAAGGTTGTGTTTTCAATAGGCAAGC 1791
Qy 540 aLeuTyrGlyTyr[illegible]yProProSerMetPro[illegible]eArgLysArgLysAspSerSerSe 560
Db 1792 ACTTTATGGCTATG[illegible]TCCAGCTTCAATGCCAA[illegible]TTTTCCCAAG---TCATCCTCCTCATC 1848
Qy 560 rCysPheSerCys[illegible]sCysProSerLysLysLysProAlaGlnAspProAlaGluValTy 580
Db 1849 TTGC---TCGTGTT[illegible]GCCCGGGCAAGAAGCAACCT---AAAGATCCATCAGAGCTTTA 1902
Qy 580 rArgAspAlaLysA[illegible]rgGluAspLeuAsnAlaAlaIlePheAsnLeuThrGluIleAspAs 600
Db 1903 TAGGGATGCAAAAC[illegible]GGAAGAACTTGATGCTG[illegible]CATCTTTAACCTTAGGGAAATTGACAA 1962
Qy 600 nTyrAspGluHis[illegible]aArgSerMetLeu[illegible]eSerGlnLeuSerPheGluLysThrPheGl 620
Db 1963 TTATGATGAGTAT[illegible]AAAGATCAATGTTGATCT[illegible]CTCAAACAAGCTTTGAGAAAACTTTTGG 2022
Qy 620 yLeuSerSerVal[illegible]eIleGluSerThrLeuMetGluAsnGlyGlyValProGluSerAl 640
Db 2023 CTTATCTT[illegible]AGTCT[illegible]CATTGAATCTACACTAATGGAGAATGGAGGAGTGGCTGAATCTGC 2082
Qy 640 aAsnSerProPro[illegible]eIleLysGluAlaIleGlnValIleGlyCysGlyTyrGluGluLy 660
Db 2083 CAACCCTTCCACAC[illegible]AATCAAGGAAGCAATT[illegible]ATGTCATCAGCTGTGGCTATGAAGAGAA 2142
Qy 660 sThrGluTrpGlyLysGlnIleGlyTrpIleTyrGlySerValThrGluAspIleLeuSe 680
Db 2143 GACTGCATGGGGGAAAGAGATTGGATGGATATATGGTTCAGTCACTGAGGATATCTTAAC 2202
Qy 680 rGlyPheLysMet[illegible]sCysArgGlyTrpArgSerIleTyrCysMetProValArgProAl 700
Db 2203 CGGCTTCAAAATG[illegible]CTGCCGAGGATGGAGATGGATTTACTGCATGCCCTTAAGGCCAGC 2262
Qy 700 aPheLysGlySerA[illegible]aProIleAsnLeuSerAspArgLeuHisGlnValLeuArgTrpAl 720
Db 2263 ATTCAAAGGATCTG[illegible]ACCCATCAATCTGTCTGATCGGTTGCACCAGGTTCTTCGATGGGC 2322
Qy 720 aLeuGlySerVal[illegible]eIlePhePheSerArg[illegible]sCysProLeuTrpTyrGlyPheGlyGl 740
Db 2323 TCTTGGAT[illegible]TGTTG[illegible]AATTTTCCTAAGCAGGCATTGCCCTCTATGGTATGGCTTTGGAGG 2382
Qy 740 yGlyArgLeuLys[illegible]rpLeuGlnArgLeuAlaTyrIleAsnThrIleValTyrProPheTh 760
Db 2383 TGGTCGTCTTAAAT[illegible]GCTTCAAAGACTAGCATATATAAACACCATTGTCTATCCTTTCAC 2442
Qy 760 rSerLeuProLeu[illegible]eAlaTyrCysThrIleProAlaValCysLeuLeuThrGlyLysPh 780
Db 2443 ATCCCTTCCACTCATTGCCTATTGTTCACTACCAGCAATCTGTCTTCTCACAGGAAAATT 2502
Qy 780 eIleIleProThrLeuSerAsnLeuAlaSerMetLeuPheLeuGlyLeuPhe[illegible]eSerIl 800
Db 2503 TATCATACCAACGCTCTCAAACCTGGCAAGTGTTCTCTTTCTTGGCCTTTTCCTTTCCAT 2562

```
Qy        800 eIleValThrAlaValLeuGluLeuArgTrpSerGlyValSerIleGluAspLeuTrpAr 820
              ||||.|||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2563 TATCGTGACTGCTGTTCTCGAGCTCCGATGGAGTGGTGTCAGCATTGAGGACTTATGGCG 2622

Qy        820 gAsnGluGlnPheTrpValIleGlyGlyValSerAlaHisLeuPheAlaValPheGlnGl 840
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2623 TAACGAGCAGTTTTGGGTCATCGGTGGCGTTTCAGCCCATCTCTTTGCCGTCTTCCAAGG 2682

Qy        840 yPheLeuLysMetLeuAlaGlyIleAspThrAsnPheThrValThrAlaLysAlaAlaGl 860
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2683 TTTCCTTAAGATGCTTGCGGGCATTGACACCAACTTTACTGTCACTGCCAAAGCAGCTGA 2742

Qy        860 uAspAlaGluPheGlyGluLeuTyrMetValLysTrpThrThrLeuLeuIleProProTh 880
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2743 TGATGCAGATTTTGGTGAGCTCTACATTGTGAAATGGACTACACTTCTAATCCCTCCAAC 2802

Qy        880 rThrLeuLeuIleIleAsnMetSerGly---CysAlaGlyPheSerAspAlaLeuAsnLy 899
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2803 AACACTCCTCATCGTCAACATGGTTGGTGTCGTTGCCGGATTCTCCGATGCCCTCAACAA 2862

Qy        899 sGlyTyrGluAlaTrpGlyProLeuPheGlyLysValPhePheAlaPheTrpValIleLe 919
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2863 AGGGTACGAAGCTTGGGGACCACTCTTTGGCAAAGTGTTCTTTTCCTTCTGGGTCATCCT 2922

Qy        919 uHisLeuTyrProPheLeuLysGlyLeuMetGlyArgGlnAsnLeuThrProThrIleVa 939
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2923 CCATCTTTATCCATTCCTCAAAGGTCTTATGGGACGCCAAAACAGGACACCAACCATTGT 2982

Qy        939 lValLeuTrpSerValLeuLeuAlaSerValPheSerLeuValTrpValLysIleAsnPr 959
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2983 TGTCCTTTGGTCAGTGTTGTTGGCTTCTGTCTTCTCTCTTGTTTGGGTTCGGATCAACCC 3042

Qy        959 oPheValAsnLysValAspAsnThrLeuValAlaGluThrCysIleSerIleAspCys 978
              |||||||:::     |||:::|'|   |||:::::::::|||||||||||||||||||
Db       3043 GTTTGTCAGCACCGCCGATAGCACCACCGTGTCACAGAGCTGCATTTCCATTGATTGT 3100

RESULT 4
AF267742
LOCUS       AF267742               2958 bp    mRNA    linear   PLN 16-FEB-2001
DEFINITION  Arabidopsis thaliana cellulose synthase catalytic subunit (IRX1)
            mRNA, complete cds.
ACCESSION   AF267742
VERSION     AF267742.1  GI:12836996
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 2958)
  AUTHORS   Taylor,N.G., Laurie,S. and Turner,S.R.
  TITLE     Multiple cellulose synthase catalytic subunits are required for
            cellulose synthesis in Arabidopsis
  JOURNAL   Plant Cell 12 (12), 2529-2540 (2000)
  MEDLINE   21066939
   PUBMED   11148295
REFERENCE   2  (bases 1 to 2958)
  AUTHORS   Taylor,N.G. and Turner,S.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-MAY-2000) Biological Sciences, University of
            Manchester, 3.614 Stopford Building, Oxford Road, Manchester M13
            9PT, UK
FEATURES             Location/Qualifiers
     source          1..2958
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
     gene            1..2958
                     /gene="IRX1"
     CDS             1..2958
                     /gene="IRX1"
                     /codon_start=1
                     /product="cellulose synthase catalytic subunit"
                     /protein_id="AAK08700.1"
                     /db_xref="GI:12836997"
                     /translation="MMESRSPICNTCGEEIGAKSNGEFFVACHECSFPICKACLEYEF
                     KEGRRICLRCGNPYDENVFDDVETKTSKTQSIVPTQTNNTSQDSGIHARHISTVSTID
                     SELNDEYGNPIWKNRVESWKDKKDKKSKKKKKDPKATKAEQQDAQVPTQQHMEDMPPN
                     TESGATDVLSVVIPIPRTKITSYRIVIIMRLIILALFFNYRITHPVDSAYGLWLTSVI
                     CEIWFAVSWVLDQFPKWSPINRETYIDRLSARFEREGEQSQLAAVDFFVSTVDPLKEP
                     PLITANTVLSILALDYPVDKVSCYVSDDGAAMLSFESLVETADFARKWVPFCKKYSIE
                     PRAPEFYFSLKIDYLRDKVQPSFVKERRAMKRDYEEFKIRMNALVAKAQKTPEEGWTM
                     QDGTSWPGNNTRDHPGMIQVFLGYSGARDIEGNELPRLVYVSREKRPGYQHHKKAGAE
                     NALVRVSAVLTNAPFILNLDCDHYVNNSKAVREAMCFLMDPVVGQDVCFVQFPQRFDG
                     IDKSDRYANRNIVFFDVNMRGLDGIQGPVYVGTGTVFRRQALYGYSPPSKPRILPQSS
                     SSSCCCLTKKKQPQDPAEIYKDAKREELDAAIFNLGDLDNYDEYDRSMLISQTSFEKT
                     FGLSTVFIESTLMENGGVPDSVNPSTLIKEAIHVISCGYEEKTEWGKEIGWIYGSITE
                     DILTGFKMHCRGWRSIYCMPLRPAFKGSAPINLSDRLHQVLRWALGSVEIFLSRHCPL
                     WYGCSGGRLKLLQRLAYINTIVYPFTSLPLVAYCTLPAICLLTGKFIIPTLSNLASML
                     FLGLFISIILTSVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQGFLKMLAGLDT
                     NFTVTSKTADDLEFGELYIVKWTTLLIPPTSLLIINLVGVVAGFSDALNKGYEAWGPL
                     FGKVFFAFWVILHLYPFLKGLMGRQNRTPTIVILWSILLASVFSLVWVRINPFVSKTD
                     TTSLSLNCLLIDC"
BASE COUNT      818 a    610 c    658 g    872 t
ORIGIN

Alignment Scores:
Pred. No.:              0              Length:          2958
Score:                  4329.50        Matches:         808
Percent Similarity:     90.51%         Conservative:    88
Best Local Similarity:  81.62%         Mismatches:      78
Query Match:            82.62%         Indels:          17
DB:                     8              Gaps:            7

US-09-980-043-2 (1-978) x AF267742 (1-2958)

Qy          1 MetMetGluSerGlyAlaProIleCysHisThrCysGlyGluGlnValGlyHisAspAla 20
              ||||||||||||  :::|||||||||:::|||||||||||:::::|||      :::
Db          1 ATGATGGAGTCTAGGTCTCCCATCTGCAACACTTGTGGTGAAGAGATTGGTGCAAAATCA 60

Qy         21 AsnGlyGluLeuPheValAlaCysHisGluCysSerTyrProMetCysLysSerCysPhe 40
              |||||||||   ||||||||||||||||||||||:::|||:::||||||:::|||
Db         61 AACGGAGAGTTCTTTGTGGCTTGTCATGAGTGTAGTTTCCCGATCTGCAAAGCTTGTCTT 120

Qy         41 GluPheGluIleAsnGluGlyArgLysValCysLeuArgCysGlySerProTyrAspGlu 60
              |||:::|||      |||||||||:::::|||||||||||||||||:::|||||||||||
Db        121 GAGTATGAATTCAAAGAAGGTCGAAGAATTTGCTTGCGTTGCGGCAATCCTTACGATGAG 180

Qy         61 AsnLeuLeuAspAspValGluLysLysGlySerGlyAsnGlnSerThrMetAlaSerHis 80
              |||:::   ||||||||||||   |||   |||      ||||||   :::   :::
Db        181 AATGTGTTTGATGATGTTGAGACAAAGACATCTAAAACTCAATCCATTGTTCCAACACAG 240

Qy         81 LeuAsnAsp---SerGlnAspValGlyIleHisAlaArgHisIleSerSerValSerThr 99
                 |||:::   ||||||||   |||||||||||||||||||||||||:::|||||||||
Db        241 ACCAATAACACTTCTCAGGATTCAGGGATTCATGCTAGACATATAAGTACAGTCTCAACA 300

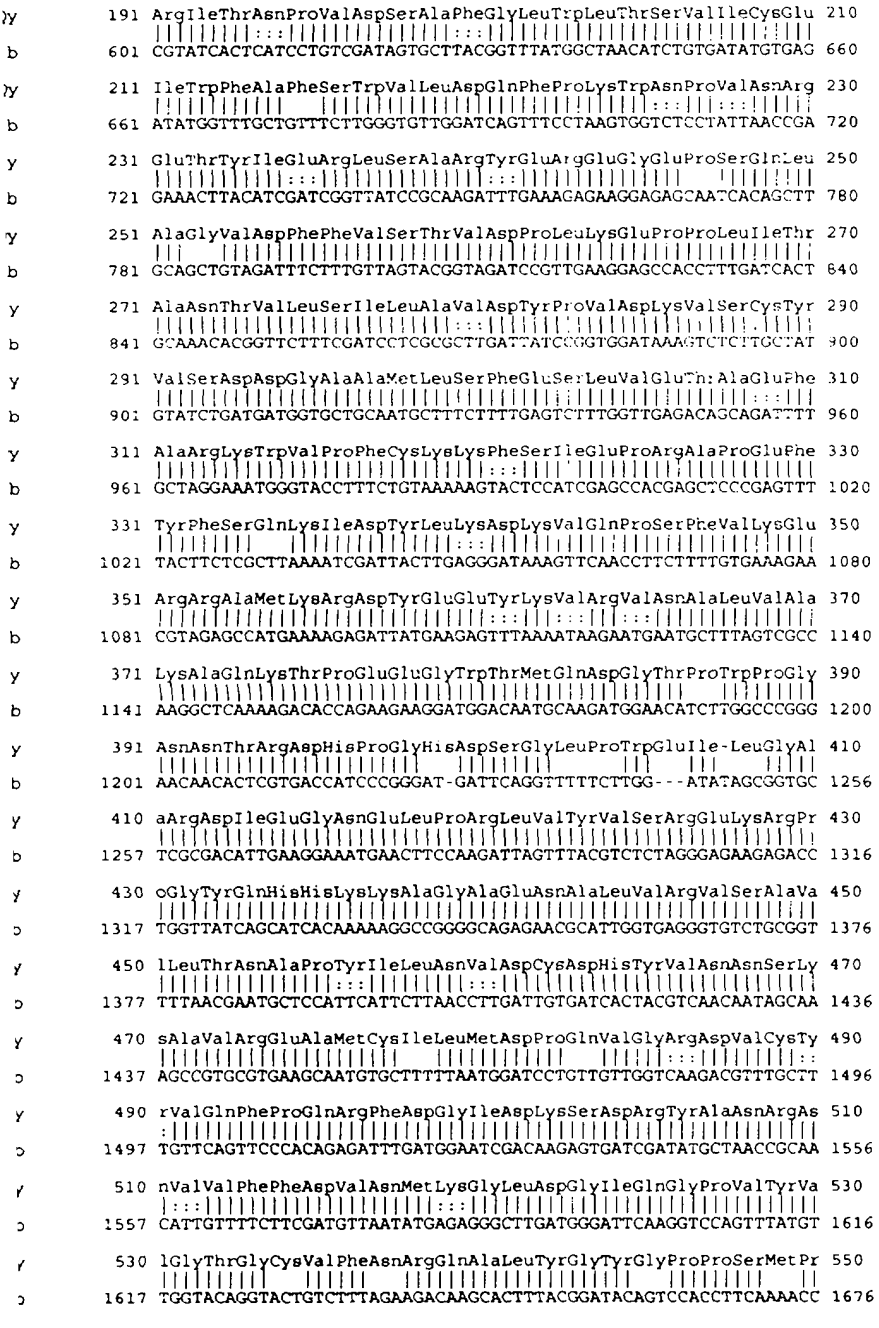
Qy        100 ValAspSerGluMetAsnAspGluTyrGlyAsnProIleTrpLysAsnArgValLysSer 119
              :::|||||||||:::|||||||||||||||||||||||||||||||||||||:::|||
Db        301 ATAGACAGTGAACTGAATGATGAATATGGCAATCCAATTTGGAAGAACAGAGTGGAGAGC 360

Qy        120 Cys---------LysAspLysGluAsnLysLysLysLysArgSerPro---------Lys 133
                         ||||||||||:::::|||||||||||||:::   |||         |||
Db        361 TGGAAAGACAAGAAAGACAAGAAGAGCAAGAAGAAGAAGAAAGATCCAAAAGCAACAAAA 420

Qy        134 AlaGluThrGluProAlaGlnValProThrGluGlnGlnMetGluGluLysPro------ 151
              ||||||   :::   |||| ||||||||||:::|||   ||||||:::   |||
Db        421 GCTGAACAACAAGACGCTCAGGTTCCTACCCAACAACACATGGAAGATATGCCACCGAAC 480

Qy        152 ---SerAlaGluAlaSerGluProLeuSerIleValTyrProIleProArgAsnLysLeu 170
                 :::   |||:::::   ||||||:::|||   ||||||||||   ||:::
Db        481 ACAGAATCTGGTGCTACAGATGTGCTTTCGGTTGTGATTCCTATCCCGAGAACAAAGATC 540

Qy        171 ThrProTyrArgAlaValIleIleMetArgLeuValIleLeuGlyLeuPhePheHisPhe 190
              :||   |||||   ||||||||||||||||||:::||||||   |||||||||::::::
Db        541 ACTTCATACAGGATCGTTATCATCATGCGGTTGATCATCTTGGCTCTGTTTTTTTAACTAC 600
```


Qy 191 ArgIleThrAsnProValAspSerAlaPheGlyLeuTrpLeuThrSerValIleCysGlu 210
Db 601 CGTATCACTCATCCTGTCGATAGTGCTTACGGTTTATGGCTAACATCTGTGATATGTGAG 660
Qy 211 IleTrpPheAlaPheSerTrpValLeuAspGlnPheProLysTrpAsnProValAsnArg 230
Db 661 ATATGGTTTGCTGTTTCTTGGGTGTTGGATCAGTTTCCTAAGTGGTCTCCTATTAACCGA 720
Qy 231 GluThrTyrIleGluArgLeuSerAlaArgTyrGluArgGluGlyGluProSerGlnLeu 250
Db 721 GAAACTTACATCGATCGGTTATCCGCAAGATTTGAAAGAGAAGGAGAGCAATCACAGCTT 780
Qy 251 AlaGlyValAspPhePheValSerThrValAspProLeuLysGluProProLeuIleThr 270
Db 781 GCAGCTGTAGATTTCTTTGTTAGTACGGTAGATCCGTTGAAGGAGCCACCTTTGATCACT 840
Qy 271 AlaAsnThrValLeuSerIleLeuAlaValAspTyrProValAspLysValSerCysTyr 290
Db 841 GCAAACACGGTTCTTTCGATCCTCGCGCTTGATTATCCGGTGGATAAAGTCTCTTGCTAT 900
Qy 291 ValSerAspAspGlyAlaAlaMetLeuSerPheGluSerLeuValGluThrAlaGluPhe 310
Db 901 GTATCTGATGATGGTGCTGCAATGCTTTCTTTTGAGTCTTTGGTTGAGACAGCAGATTTT 960
Qy 311 AlaArgLysTrpValProPheCysLysLysPheSerIleGluProArgAlaProGluPhe 330
Db 961 GCTAGGAAATGGGTACCTTTCTGTAAAAAGTACTCCATCGAGCCACGAGCTCCCGAGTTT 1020
Qy 331 TyrPheSerGlnLysIleAspTyrLeuLysAspLysValGlnProSerPheValLysGlu 350
Db 1021 TACTTCTCGCTTAAAATCGATTACTTGAGGGATAAAGTTCAACCTTCTTTTGTGAAAGAA 1080
Qy 351 ArgArgAlaMetLysArgAspTyrGluGluTyrLysValArgValAsnAlaLeuValAla 370
Db 1081 CGTAGAGCCATGAAAAGAGATTATGAAGAGTTTAAAATAAGAATGAATGCTTTAGTCGCC 1140
Qy 371 LysAlaGlnLysThrProGluGluGlyTrpThrMetGlnAspGlyThrProTrpProGly 390
Db 1141 AAGGCTCAAAAGACACCAGAAGAAGGATGGACAATGCAAGATGGAACATCTTGGCCCGGG 1200
Qy 391 AsnAsnThrArgAspHisProGlyHisAspSerGlyLeuProTrpGluIle-LeuGlyAl 410
Db 1201 AACAACACTCGTGACCATCCCGGGAT-GATTCAGGTTTTTCTTGG---ATATAGCGGTGC 1256
Qy 410 aArgAspIleGluGlyAsnGluLeuProArgLeuValTyrValSerArgGluLysArgPr 430
Db 1257 TCGCGACATTGAAGGAAATGAACTTCCAAGATTAGTTTACGTCTCTAGGGAGAAGAGACC 1316
Qy 430 oGlyTyrGlnHisHisLysLysAlaGlyAlaGluAsnAlaLeuValArgValSerAlaVa 450
Db 1317 TGGTTATCAGCATCACAAAAAGGCCGGGGCAGAGAACGCATTGGTGAGGGTGTCTGCGGT 1376
Qy 450 lLeuThrAsnAlaProTyrIleLeuAsnValAspCysAspHisTyrValAsnAsnSerLy 470
Db 1377 TTTAACGAATGCTCCATTCATTCTTAACCTTGATTGTGATCACTACGTCAACAATAGCAA 1436
Qy 470 sAlaValArgGluAlaMetCysIleLeuMetAspProGlnValGlyArgAspValCysTy 490
Db 1437 AGCCGTGCGTGAAGCAATGTGCTTTTTAATGGATCCTGTTGTTGGTCAAGACGTTTGCTT 1496
Qy 490 rValGlnProProGlnArgPheAspGlyIleAspLysSerAspArgTyrAlaAsnArgAs 510
Db 1497 TGTTCAGTTCCCACAGAGATTTGATGGAATCGACAAGAGTGATCGATATGCTAACCGCAA 1556
Qy 510 nValValPhePheAspValAsnMetLysGlyLeuAspGlyIleGlnGlyProValTyrVa 530
Db 1557 CATTGTTTTCTTCGATGTTAATATGAGAGGGCTTGATGGGATTCAAGGTCCAGTTTATGT 1616
Qy 530 lGlyThrGlyCysValPheAsnArgGlnAlaLeuTyrGlyTyrGlyProProSerMetPr 550
Db 1617 TGGTACAGGTACTGTCTTTAGAAGACAAGCACTTTACGGATACAGTCCACCTTCAAAACC 1676
Qy 550 oSerLeuArgLysArgLysAspSerSerSerCysPheSerCysCysCysProSerLysLy 570
Db 1677 GAGGATTTTACCGAAATCTTCATCATCG --- --- TCGTGTTGCTGTCTAACCAAGAA 1727
Qy 570 sLysProAlaGlnAspProAlaGluValTyrArgAspAlaLysArgGluAspLeuAsnAl 590
Db 1728 GAAACAACCCCAAGATGCTGCCCAGATTTACAAAGACGCAAAGCGAGAAGAACTTGATGC 1787
Qy 590 aAlaIlePheAsnLeuThrGluIleAspAsnTyrAspGluHisGluArgSerMetLeuIl 610
Db 1788 AGCAATCTTTAATCTTGGGGACCTAGACAACTACGATGAATACGACAGATCAATGCTGAT 1847
Qy 610 eSerGlnLeuSerPheGluLysThrPheGlyLeuSerSerValPheIleGluSerThrLe 630
Db 1848 TTCACAAACAAGCTTTGAGAAAACGTTTGGTCTCTCTACGGTTTTCATCGAGTCTACTCT 1907
Qy 630 uMetGluAsnGlyIleValProGluSerAlaAsnSerProProPheIleLysGluAlaIl 650
Db 1908 TATGGAAAATGTAGTTCCTGAATCAGCAAATAGTCCTCCACCTTTCATCAAAGAAGCTAT 1967
Qy 650 eGlnValPheGly[illegible]GluTrpGlyLysGlnIleGlyTrpIl 670
Db 1968 TCATGTCATTAGTTTGGAAACGAAAGAGAAAATCGAATGGGGAAAAGAAATCGGATGGAT 2027
Qy 670 eTyrGlySerValThrGluAspIleLeuSerGlyPheLysMetHisCysArgGlyTrpAr 690
Db 2028 TTACGGGTCGATCATGGAAGACATTTTGACGGGTTTCAAGATGCATTGTCGCGGATGGAG 2087
Qy 690 gSerIleTyrCysMetProValArgProAlaPheLysGlySerAlaProIleAsnLeuSe 710
Db 2088 GTCGATTTACTGTATGCCATTAAGACCAGCTTTTAAAGGTTCTGCACCAATCAATCTATC 2147
Qy 710 rAspArgLeuHisGlnValLeuArgTrpAlaLeuGlySerValGluIlePhePheSerAr 730
Db 2148 AGATAGGCTTCACCAGGTTCTACGTTGGGCTCTTGGCTCGGTTGAGATCTTCCTTAGCCG 2207
Qy 730 gHisCysProLeuTrpTyrGlyPheGlyGlyGlyArgLeuLysTrpLeuGlnArgLeuAl 750
Db 2208 ACATTGTCCTTTGTGGTACGGTTGCAGCGGAGGCCGCCTCAAGTTACTCCAGAGATTAGC 2267
Qy 750 aTyrIleAsnThrIleValTyrProPheThrSerLeuProLeuIleAlaTyrCysThrIl 770
Db 2268 TTATATAAACACTATTGTCTACGCATTCACTTCTTTGCCTCTTGTTGCTTACTGTACTCT 2327
Qy 770 eProAlaValCysLeuLeuThrGlyLysPheIleIleProThrLeuSerAsnLeuAlaSe 790
Db 2328 TCCAGCTATTTGCCTTCTTACCGGCAAATTTATCATCCCAACGCTATCAAACCTAGCAAG 2387
Qy 790 rMetLeuPheLeuGlyLeuPheIleSerIleIleValThrAlaValLeuGluLeuArgTr 810
Db 2388 CATGCTGTTTCTAGGTCTCTTTATATCAATCATCTTAACGAGTGTCCTCGAGCTTCGATG 2447
Qy 810 pSerGlyValSerIleGluAspLeuTrpArgAsnGluGlnPheTrpValIleGlyGlyVa 830
Db 2448 GAGCGGAGTCAGTATCGAAGACTTATGGAGAAACGAACAGTTTTGGGTTATTGGAGGTGT 2507
Qy 830 lSerAlaHisLeuPheAlaValPheGlnGlyPheLeuLysMetLeuAlaGlyIleAspTh 850
Db 2508 CTCAGCTCATCTCTTTGCCGTTTTCCAAGGATTCCTCAAAATGCTAGCTGGTCTCGACAC 2567
Qy 850 rAsnPheThrValThrAlaLysAlaAlaGluAspAlaGluPheGlyGluLeuTyrMetVa 870
Db 2568 AAATTTCACAGTCACATCGAAAACCGCAGATGATTTAGAATTCGGTGAGCTTTACATCGT 2627
Qy 870 lLysTrpThrThrLeuLeuIleProProThrThrLeuLeuIleIleAsnMetSerGly-- 889
Db 2628 CAAAATGGACAACTCTCTTGATCCCTCCGACTTCACTTCTTATAATTAACTTGGTCGGAGT 2687
Qy 890 -CysAlaGlyPheSerAspAlaLeuAsnLysGlyTyrGluAlaTrpGlyProLeuPheGl 909
Db 2688 CGTCGCTGGATTCTCTGATGCTCTTAACAAAGGTTATGAAGCTTGGGGACCTTTGTTTGG 2747
Qy 909 yLysValPhePheAlaPheTrpValIleLeuHisLeuTyrProPheLeuLysGlyLeuMe 929

```
Db       2748 GAAGGTCTTTTTCGCCTTTTGGGTGATTCTTCATCTTTATCCATTCCTCAAAGGTCTTAT 2807

Qy        929 tGlyArgGlnAsnLeuThrProThrIleValValLeuTrpSerValLeuLeuAlaSerVa 949

Db       2808 GGGAAGACAAAACAGAACACCTACTATTGTTATTCTCTGGTCTATCTTGCTTGCTTCTGT 2867

Qy        949 lPheSerLeuValTrpValLysIleAsnProPheValAsnLysValAspAsnThrLeuVa 969

Db       2868 GTTTTCACTTGTTTGGGTTCGTATCAACCCTTTCGTCTCCAAAACCGATACGACTTCCCT 2927

Qy        969 lAlaGluThrCysIleSerIleAspCys 978

Db       2928 TTCTCTGAACTGTCTTTTGATCGATTGC 2955

RESULT 5
AY099636
LOCUS       AY099636               3109 bp    mRNA    linear   PLN 06-MAY-2002
DEFINITION  Arabidopsis thaliana cellulose synthase-like protein (At4g18780)
            mRNA, complete cds.
ACCESSION   AY099636
VERSION     AY099636.1  GI:20466339
KEYWORDS    FLI_CDNA.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 3109)
  AUTHORS   Nguyen,M., Karlin-Neumann,G., Southwick,A., Lam,B., Miranda,M.,
            Palm,C.J., Bowser,L., Jones,T., Banh,J., Carninci,P., Chen,H.,
            Cheuk,R., Chung,M.K., Hayashizaki,Y., Ishida,J., Kamiya,A.,
            Kawai,J., Kim,C., Lin,J., Liu,S.X., Narusaka,M., Pham,P.K.,
            Sakano,H., Sakurai,T., Satou,M., Seki,M., Shinn,P., Yamada,K.,
            Shinozaki,K., Ecker,J., Theologis,A. and Davis,R.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-APR-2002) DNA Sequencing and Technology Center,
            Stanford University, 855 California Avenue, Palo Alto, CA 94304,
            USA
COMMENT     e-mail for correspondence: arabasequence.stanford.edu

            RIKEN Genomic Sciences Center (GSC) members carried out the
            collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN
            Arabidopsis  Full-Length cDNA'): Seki,M., Narusaka,M., Ishida,J.,
            Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J.,
            Hayashizaki,Y. and Shinozaki,K.

            The Salk, Stanford, PGSC (SSP) Consortium members carried out the
            sequencing and annotation of the RAFL cDNAs: Nguyen,M.,
            Southwick,A., Karlin-Neumann,G., Lam,B., Miranda,M., Palm,C.J.,
            Bowser,L., Jones,T., Banh,J., Chen,H., Cheuk,R., Chung,M.K.,
            Kim,C., Lin,J., Liu,S.X., Pham,P.K., Sakano,H., Shinn,P.,
            Yamada,K., Ecker,J., Theologis,A. and Davis,R.W.

            Nguyen,M. (SSP/Stanford) and Seki,M. (RIKEN GSC) contributed
            equally to this work.  Shinozaki,K. (RIKEN GSC) and Davis,R.W.
            (SSP/Stanford)  contributed equally to this work as PIs.
FEATURES             Location/Qualifiers
     source          1. .3109
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
                     /chromosome="4"
                     /clone="RAFL09-65-M12 (R24904)"
                     /note="This clone is in pBluescrip"
                     ecotype: Columbia"
     gene            1. .3109
                     /gene="At4g18780"
                     /note="synonym: F28A21.190"
     CDS             68. .3025
                     /gene="At4g18780"
                     /codon_start=1
                     /product="cellulose synthase-like protein"
                     /protein_id="AAM20487.1"
                     /db_xref="GI:20466340"
                     /translation="MMESRSPICNTCGEEIGVKSNGEFFVACHECSFPICKACLEYEF
                     KEGRRICLRCGNPYDENVFDDVETKTSKTQSIVPTQTNNTSQDSGIHARHISTVSTID
                     SELNDEYGNPIWKNRVESWKDKKDKKSKKKKKDPKATKAEQHEAQIPTQQHMEDTPPN
                     TESGATDVLSVVIPIPRTKITSYRIVIIMRLIILALFFNYRITHPVDSAYGLWLTSVI
                     CEIWFAVSWVLDQFPKWSPINRETYIDRLSARFEREGEQSQLAAVDFFVSTVDPLKEP
                     PLITANTVLSILALDYPVDKVSCYVSDDGAAMLSFESLVETADFARKWVPFCKKYSIE
                     PRAPEFYFSLKIDYLRDKVQPSFVKERRAMKRDYEEFKIRMNALVAKAQKTPEEGWTM
                     QDGTSWPGNNTRDHPGMIQVFLGYSGARDIEGNELPRLVYVSREKRPGYQHHKKAGAE
                     NALVRVSAVLTNAPFILNLDCDHYVNNSKAVREAMCFLMDPVVGQDVCFVQFPQRFDG
                     IDKSDRYANRNIVFFDVNMRGLDGIQGPVYVGTGTVFRRQALYGYSPPSKPRILPQSS
                     SSSCCCLTKKKQPQDPSEIYKDAKREELDAAIFNLGDLDNYDEYDRSMLISQTSFEKT
                     FGLSTVFIESTLMENGGVPDSVNPSTLIKEAIHVISCGYEEKTEWGKEIGWIYGSITE
                     DILTGFKMHCRGWRSIYCMPLRPAFKGSAPINLSDRLHQVLRWALGSVEIFLSRHCPL
                     WYGCSGGRLKLLQRLAYINTIVYPFTSLPLVAYCTLPAICLLTGKFIIPTLSNLASML
                     FLGLFISIILTSVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQGFLKMLAGLDT
                     NFTVTSKTADDLEFGELYIVKWTTLLIPPTSLLIINLVGVVAGFSDALNKGYEAWGPL
                     FGKVFFAFWVILHLYPFLKGLMGRQNRTPTIVILWSILLASVFSLVWVRINPFVSKTD
                     TTSLSLNCLLIDC"
BASE COUNT      859 a    631 c    687 g    932 t
ORIGIN

Alignment Scores:
Pred. No.:                0              Length:          3109
Score:                    4322.50        Matches:         806
Percent Similarity:       90.40%         Conservative:    89
Best Local Similarity:    81.41%         Mismatches:      79
Query Match:              82.49%         Indels:          17
DB:                       8              Gaps:            7

US-09-980-043-2 (1-978) x AY099636 (1-3109)

Qy          1 MetMetGluSerGlyAlaProIleCysHisThrCysGlyGluGlnValGlyHisAspAla 20

Db         68 ATGATGGAGTCTAGGTCTCCCATCTGCAACACTTGTGGTGAAGAGATTGGTGTAAAATCA 127

Qy         21 AsnGlyGluLeuPheValAlaCysHisGluCysSerTyrProMetCysLysSerCysPhe 40

Db        128 AACGGAGAGTTCTTTGTGGCTTGTCATGAGTGTAGTTTCCCAATCTGCAAAGCTTGTCTT 187

Qy         41 GluPheGluIleAsnGluGlyArgLysValCysLeuArgCysGlySerProTyrAspGlu 60

Db        188 GAGTATGAATTCAAAGAAGGTCGAAGAATTTGCTTGCGTTGCGGCAATCCTTACGATGAG 247

Qy         61 AsnLeuLeuAspAspValGluLysLysGlySerGlyAsnGlnSerThrMetAlaSerHis 80

Db        248 AATGTGTTTGATGATGTTGAGACAAAGACATCTAAAACTCAATCCATTGTTCCAACACAG 307

Qy         81 LeuAsnAsp---SerGlnAspValGlyIleHisAlaArgHisIleSerSerValSerThr 99

Db        308 ACCAATAACACTTCTCAGGATTCAGGGATTCATGCTAGACATATAAGTACAGTCTCAACA 367

Qy        100 ValAspSerGluMetAsnAspGluTyrGlyAsnProIleTrpLysAsnArgValLysSer 119

Db        368 ATAGACAGTGAACTGAATGATGAATATGGCAATCCAATTTGGAAGAACAGAGTGGAGAGC 427

Qy        120 Cys---------LysAspLysGluAsnLysLysLysLysArgSerPro---------Lys 133

Db        428 TGGAAAGACAAGAAAGACAAGAAGAGCAAGAAGAAGAAGAAAGATCCAAAAGCAACAAAA 487

Qy        134 AlaGluThrGluProAlaGlnValProThrGluGlnGlnMetGluGluLysPro------ 151

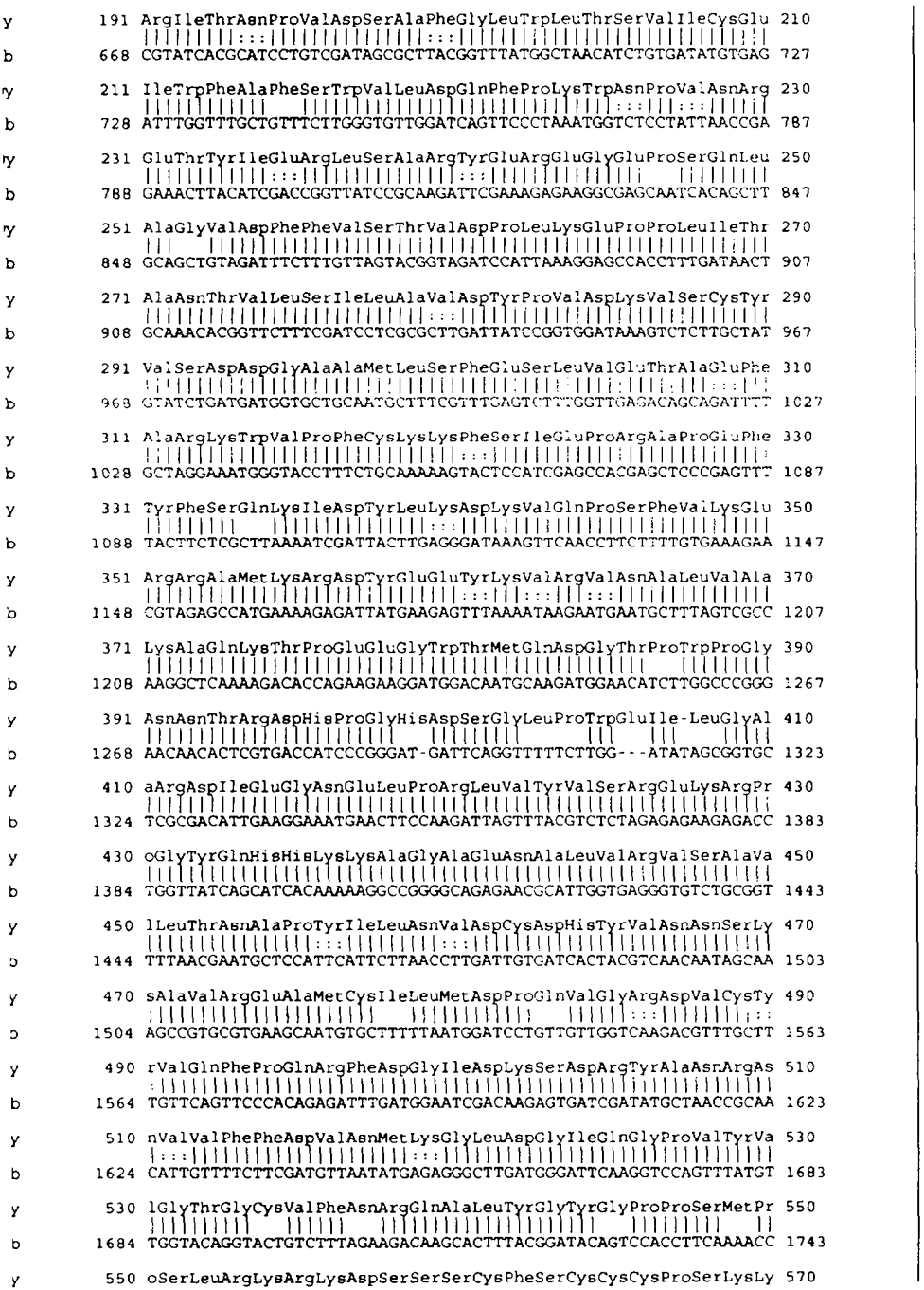
Db        488 GCTGAACAACATGAGGCTCAGATTCCTACCCAACAGCACATGGAAGATACGCCACCGAAC 547

Qy        152 ---SerAlaGluAlaSerGluProLeuSerIleValTyrProIleProArgAsnLysLeu 170

Db        548 ACAGAATCTGGTGCTACAGATGTGCTTTCGGTTGTGATTCCTATCCCAAGGACAAAAATC 607

Qy        171 ThrProTyrArgAlaValIleIleMetArgLeuValIleLeuGlyLeuPhePheHisPhe 190

Db        608 ACTTCATATAGGATTGTCATCATCATGCGGTTGATCATCTTGGCTCTGTTTTTTAACTAC 667
```


y 191 ArgIleThrAsnProValAspSerAlaPheGlyLeuTrpLeuThrSerValIleCysGlu 210
b 668 CGTATCACGCATCCTGTCGATAGCGCTTACGGTTTATGGCTAACATCTGTGATATGTGAG 727
y 211 IleTrpPheAlaPheSerTrpValLeuAspGlnPheProLysTrpAsnProValAsnArg 230
b 728 ATTTGGTTTGCTGTTTCTTGGGTGTTGGATCAGTTCCCTAAATGGTCTCCTATTAACCGA 787
y 231 GluThrTyrIleGluArgLeuSerAlaArgTyrGluArgGluGlyGluProSerGlnLeu 250
b 788 GAAACTTACATCGACCGGTTATCCGCAAGATTCGAAAGAGAAGGCGAGCAATCACAGCTT 847
y 251 AlaGlyValAspPhePheValSerThrValAspProLeuLysGluProProLeuIleThr 270
b 848 GCAGCTGTAGATTTCTTTGTTAGTACGGTAGATCCATTAAAGGAGCCACCTTTGATAACT 907
y 271 AlaAsnThrValLeuSerIleLeuAlaValAspTyrProValAspLysValSerCysTyr 290
b 908 GCAAACACGGTTCTTTCGATCCTCGCGCTTGATTATCCGGTGGATAAAGTCTCTTGCTAT 967
y 291 ValSerAspAspGlyAlaAlaMetLeuSerPheGluSerLeuValGluThrAlaGluPhe 310
b 968 GTATCTGATGATGGTGCTGCAATGCTTTCGTTTGAGTCTTTGGTTGAGACAGCAGATTTT 1027
y 311 AlaArgLysTrpValProPheCysLysLysPheSerIleGluProArgAlaProGluPhe 330
b 1028 GCTAGGAAATGGGTACCTTTCTGCAAAAAGTACTCCATCGAGCCACGAGCTCCCGAGTTT 1087
y 331 TyrPheSerGlnLysIleAspTyrLeuLysAspLysValGlnProSerPheValLysGlu 350
b 1088 TACTTCTCGCTTAAAATCGATTACTTGAGGGATAAAGTTCAACCTTCTTTTGTGAAAGAA 1147
y 351 ArgArgAlaMetLysArgAspTyrGluGluTyrLysValArgValAsnAlaLeuValAla 370
b 1148 CGTAGAGCCATGAAAAGAGATTATGAAGAGTTTAAAATAAGAATGAATGCTTTAGTCGCC 1207
y 371 LysAlaGlnLysThrProGluGluGlyTrpThrMetGlnAspGlyThrProTrpProGly 390
b 1208 AAGGCTCAAAAGACACCAGAAGAAGGATGGACAATGCAAGATGGAACATCTTGGCCCGGG 1267
y 391 AsnAsnThrArgAspHisProGlyHisAspSerGlyLeuProTrpGluIle-LeuGlyAl 410
b 1268 AACAACACTCGTGACCATCCCGGGAT-GATTCAGGTTTTTCTTGG---ATATAGCGGTGC 1323
y 410 aArgAspIleGluGlyAsnGluLeuProArgLeuValTyrValSerArgGluLysArgPr 430
b 1324 TCGCGACATTGAAGGAAATGAACTTCCAAGATTAGTTTACGTCTCTAGAGAGAAGAGACC 1383
y 430 oGlyTyrGlnHisHisLysLysAlaGlyAlaGluAsnAlaLeuValArgValSerAlaVa 450
b 1384 TGGTTATCAGCATCACAAAAAGGCCGGGGCAGAGAACGCATTGGTGAGGGTGTCTGCGGT 1443
y 450 lLeuThrAsnAlaProTyrIleLeuAsnValAspCysAspHisTyrValAsnAsnSerLy 470
b 1444 TTTAACGAATGCTCCATTCATTCTTAACCTTGATTGTGATCACTACGTCAACAATAGCAA 1503
y 470 sAlaValArgGluAlaMetCysIleLeuMetAspProGlnValGlyArgAspValCysTy 490
b 1504 AGCCGTGCGTGAAGCAATGTGCTTTTTAATGGATCCTGTTGTTGGTCAAGACGTTTGCTT 1563
y 490 rValGlnPheProGlnArgPheAspGlyIleAspLysSerAspArgTyrAlaAsnArgAs 510
b 1564 TGTTCAGTTCCCACAGAGATTTGATGGAATCGACAAGAGTGATCGATATGCTAACCGCAA 1623
y 510 nValValPhePheAspValAsnMetLysGlyLeuAspGlyIleGlnGlyProValTyrVa 530
b 1624 CATTGTTTTCTTCGATGTTAATATGAGAGGGCTTGATGGGATTCAAGGTCCAGTTTATGT 1683
y 530 lGlyThrGlyCysValPheAsnArgGlnAlaLeuTyrGlyTyrGlyProProSerMetPr 550
b 1684 TGGTACAGGTACTGTCTTTAGAAGACAAGCACTTTACGGATACAGTCCACCTTCAAAACC 1743
y 550 oSerLeuArgLysArgLysAspSerSerSerCysPheSerCysCysCysProSerLysLy 570

Db 1744 GAGGATTTTACCG AATCTTCATCATG---- - --TCGTGTTGCTGTCTAACCAAGAA 1794
Qy 570 sLysProAlaGlnAspProAlaGluValTyrArgAspAlaLysArgGluAspLeuAsnAl 590
Db 1795 GAAACAACCTCAAGATCCTTCCGAGATTTACAAAGATGCAAAGCGAGAAGAACTTGATGC 1854
Qy 590 aAlaIlePheAsnLeuThrGluIleAspAsnTyrAspGluHisGluArgSerMetLeuIl 610
Db 1855 TGCAATCTTTAATCTTGGGGACCTAGACAACTACGATGAGTACGACAGATCGATGCTGAT 1914
Qy 610 eSerGlnLeuSerPheGluLysThrPheGlyLeuSerSerValPheIleGluSerThrLe 630
Db 1915 TTCACAAACAAGCTTTGAGAAAACGTTTGGTCTCTCTACGGTTTTCATCGAGTCTACTCT 1974
Qy 630 uMetGluAsnGlyGlyValProGluSerAlaAsnSerProProPheIleLysGluAlaIl 650
Db 1975 TATGGAGAATGGCGGTGTTCCTGACTCTGTAAACCCGTCAACGCTCATCAAAGAAGCTAT 2034
Qy 650 eGlnValIleGlyLysGlyTyrGluGluLysThrGluTrpGlyLysGlnIleGlyTrpIl 670
Db 2035 TCATGT AT AG T ATA AA A AAAA TGAATGGGGAAAAGAAATACGATGGAT 2094
Qy 670 eTyrG Val lu sp eu yPheLysMetHisCysArgGlyTrpAr 690
Db 2095 TTACGGGTC ATCAC GAAGA ATTTTGACGGTTTTCAAGATGCATTGTCGTGGATGGAG 2154
Qy 690 gSerIleTyrCysMetProValArgProAlaPheLysGlySerAlaProIleAsnLeuSe 710
Db 2155 GTCGATTTACTGTATGCCATTAAGACCAGCATTTAAAGGATCTGCTCCAATCAATCTATC 2214
Qy 710 rAspArgLeuHisGlnValLeuArgTrpAlaLeuGlySerValGluIlePhePheSerAr 730
Db 2215 AGATAGGCTTCACCAGGTTCTACGTTGGGCTCTTGGCTCGGTTGAGATCTTCCTTAGCCG 2274
Qy 730 gHisCysProLeuTrpTyrGlyPheGlyGlyArgLeuLysTrpLeuGlnArgLeuAl 750
Db 2275 ACATTGTCCTTTGTGGTACGGTTGCAGCGGAGGCCGCCTCAAGTTGCTCCAGAGATTAGC 2334
Qy 750 aTyrIleAsnThrIleValTyrProPheThrSerLeuProLeuIleAlaTyrCysThrIl 770
Db 2335 TTATATAAACACTATTGTCTACCCATTCACTTCTTTGCCTCTTGTTGCTTACTGTACTCT 2394
Qy 770 eProAlaValCysLeuLeuThrGlyLysPheIleIleProThrLeuSerAsnLeuAlaSe 790
Db 2395 TCCAGCTATTTGCCTTCTTACCGGCAAATTTATCATCCCAACGCTATCAAACCTAGCAAG 2454
Qy 790 rMetLeuPheLeuGlyLeuPheIleSerIleIleValThrAlaValLeuGluLeuArgTr 810
Db 2455 CATGCTGTTTCTAGGTCTCTTTATATCAATCATCTTAACGAGTGTCCTCGAGCTTCGATG 2514
Qy 810 pSerGlyValSerIleGluAspLeuTrpArgAsnGluGlnPheTrpValIleGlyGlyVa 830
Db 2515 GAGCGGAGTCAGTATCGAAGACTTATGGAGAAACGAACAGTTTTGGGTTATTGGAGGTGT 2574
Qy 830 lSerAlaHisLeuPheAlaValPheGlnGlyPheLeuLysMetLeuAlaGlyIleAspTh 850
Db 2575 CTCAGCTCATCTCTTTGCCGTTTTCCAAGGATTCCTCAAAATGCTCGCTGGTCTCGACAC 2634
Qy 850 rAsnPheThrValThrAlaLysAlaAlaGluAspAlaGluPheGlyGluLeuTyrMetVa 870
Db 2635 AAATTTCACAGTCACATCGAAAACCGCAGATGATTTAGAATTCGGTGAGCTTTACATTGT 2694
Qy 870 lLysTrpThrThrLeuLeuIleProProThrThrLeuLeuIleIleAsnMetSerGly- 889
Db 2695 CAAATGGACAACTCTCTTGATCCCTCCGACGTCACTTCTTATAATCAACTTGGTCGGAGT 2754
Qy 890 -CysAlaGlyPheSerAspAlaLeuAsnLysGlyTyrGluAlaTrpGlyProLeuPheGl 909
Db 2755 TGTTGCTGGATTCTCTGACGCTCTTAACAAAGGTTATGAAGCTTGGGGACCTTTGTTTGG 2814
Qy 909 yLysValPhePheAlaPheTrpValIleLeuHisLeuTyrProPheLeuLysGlyLeuMe 929

```
Db       2815 GAAGGTATTTTTCGCCTTTTGGGTGATTCTTCATCTTTATGCATTCGTCAAAGGTCTTAT 2874

Qy        929 tGlyArgGlnAsnLeuThrProThrIleValValLeuTrpSerValLeuLeuAlaSerVa 949

Db       2875 GGGAAGACAAAACAGAACACCTACTATTGTTATTCTCTGGTCTATCTTGCTTGCTTCTGT 2934

Qy        949 lPheSerLeuValTrpValLysIleAsnProPheValAsnLysValAspAsnThrLeuVa 969

Db       2935 GTTTTCACTTGTTTGGGTTCGTATCAATCCTTTCGTCTCCAAAACCCATACGACTTCCCT 2994

Qy        969 lAlaGluThrCysIleSerIleAspCys 978

Db       2995 TTCTCTGAACTGTCTTTTGATCGATTGC 3022

RESULT 6
AF323039
LOCUS       AF323039               3145 bp    mRNA    linear   PLN 25-FEB-2002
DEFINITION  Zinnia elegans cellulose synthase CesA 1 mRNA, complete cds.
ACCESSION   AF323039
VERSION     AF323039.2  GI:18875453
KEYWORDS    .
SOURCE      Zinnia elegans
  ORGANISM  Zinnia elegans
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; campanulids; Asterales; Asteraceae; Asteroideae;
            Heliantheae; Zinnia.
REFERENCE   1  (bases 1 to 3145)
  AUTHORS   Haigler,C.H., Ivanova Datcheva,M., Hogan,P.S., Salnikov,V.V.,
            Hwang,S., Martin,K. and Delmer,D.P.
  TITLE     Carbon partitioning to cellulose synthesis
  JOURNAL   Plant Mol. Biol. 47 (1-2), 29-51 (2001)
  MEDLINE   21438255
   PUBMED   11554477
REFERENCE   2  (bases 1 to 3145)
  AUTHORS   Hwang,S. and Haigler,C.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-NOV-2000) Biological Sciences, Texas Tech University,
            Box 43131, Lubbock, TX 79409, USA
REFERENCE   3  (bases 1 to 3145)
  AUTHORS   Hwang,S. and Haigler,C.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-FEB-2002) Biological Sciences, Texas Tech University,
            Box 43131, Lubbock, TX 79409, USA
  REMARK    Sequence update by submitter
COMMENT     On Feb 25, 2002 this sequence version replaced gi:13021930.
FEATURES             Location/Qualifiers
     source          1..3145
                     /organism="Zinnia [illegible]"
                     /mol_type="mRNA"
                     /db_xref="taxon:[illegible]"
     CDS             69..3005
                     /codon_start=1
                     /product="cellulose synthase CesA 1"
                     /protein_id="AAK11568.2"
                     /db_xref="GI:18875454"
                     /translation="MMHPETPICNTCGEHIGPGNNGNGVFVGCHECNFPICKSCLEYE
                     INEGHTACIQCGTPYDGNLTKVGEMEKEPVTRNTMAAHLSNSQDAGLHARNVSTVSTV
                     DTEMTDESGNPIWKNRVESWKDKKNKKKKGGDKVAKEVQVPEDQHIEEKQQSADPNAM
                     QPPSQIIPIPKSQITPYRIVIILRLIILGLFFHYRITNPVESSYGLWLTSVICEIWFA
                     ISWVLDQFPKWYPINRVTFTDELSARYEREGEPSKLAAVDFFVSTVDPLKEPPLITAN
                     TVLSILAVDYPVEKVSCYVSDDGAAMLTFESLVETAEFARKWVPFCKKFSIEPRAPEF
                     YFSQKIDYLKDKVQPSFVKERRAMKRDYEEYKVRVNALVAKAQKTPEEGWTMQDGTPW
                     PGNNPRDHPGMIQVFLGHSGAHDIEGNELPRLVYVSREKRPGYQHHKKAGAENALVRV
                     SAVLTNAPYILNLDCDHYVNNSKAVREAMCFMMDPQVGRDVCYIQFPQRFDGIDRSDR
                     YANRNTVFFDVNMKGLDGIQGPVYVGTGCVFYRQALYGYDPQSLPTLPSPSSSSSCCC
                     CGPKKPKKDLEEFKRDARRDELNAAIFNLKEIESYDYESLLISQMSFEKTFGMSSV
                     FIESTLMENGGLAESANPATMINEAIHVISCGYEDKTAWGKEIGWIYGSVTEDILTGF
                     KMHCRGWRSIYCMPVRPAFKGSAPINLSDRLHQVLRWALGSVEIFLSRHCPLWYGWGG
                     GRLKLLQRLAYINTIVYPFTSLPLVAYCTLPAICLLTGKFIIPTLSNIAAVWFLGLFL
                     SIITTSVLEIRWSGVSIEELWRNEQFWVIGGVSAHLFAVFQGFLKMLAGVDTNFTVTA
                     KAADDQEFGELYMIKWTTVLIPPTTLLVLNLVGVVAGFSFALNKGYEAWGPLFGKVFF
                     AFWVILHLYPFLKGLMGRQNRTPTIVILWSVLLASVFSLVWVRIDPFVSKGDSNLTQG
```

```
                     CIAIDC"
BASE COUNT      873 a    582 c    746 g    944 t
ORIGIN

Alignment Scores:
Pred. No.:                0          Length:          3145
Score:                    4220.00    Matches:         790
Percent Similarity:       89.35%     Conservative:    91
Best Local Similarity:    80.12%     Mismatches:      90
Query Match:              80.53%     Indels:          18
DB:                       8          Gaps:            10

US-09-980-043-2 (1-978) x AF323039 (1-3145)

Qy          1 MetMetGluSerGlyAlaProIleCysHisThrCysGlyGluGlnValGlyHis------ 18

Db         69 ATGATGCATCCAGAGACTCCCATTTGCAACACTTGTGGGGAGCATATTGGGTTTGGTCCT 128

Qy         19 AspAlaAsnGlyGluLeuPheValAlaCysHisGluCysSerTyrProMetCysLysSer 38

Db        129 GATGGGAATGGA---GTGTTTGTGGGTTGTCATGAATGTAATTTTCCTATTTGCAAGTCA 185

Qy         39 CysPheGluPheGluIleAsnGluGlyArgLysValCysLeuArgCysGlySerProTyr 58

Db        186 TGTTTGGAGTATGAGATCAATGAGGGTCACACTGCTTGCATTCAGTGTGGTACACCTTAT 245

Qy         59 AspGluAsnLeu------LeuAspAspValGluLysLysGlySerGlyAsnGlnSerThr 76

Db        246 GATGGGAATTTGACAAAGGTTGGTGAGATGGAAAAGGAG---CCTGTAACCCGGAATACT 302

Qy         77 MetAlaSerHisLeuAsnAspSerGlnAspValGlyIleHisAlaArgHisIleSerSer 96

Db        303 ATGGCTGCACACCTTAGTAACAGTCAGGATGCTGGACTCCATGCGAGAAATGTAAGCACC 362

Qy         97 ValSerThrValAspSerGluMetAsnAspGluTyrGlyAsnProIleTrpLysAsnArg 116

Db        363 GTATCAACAGTCGATACTGAGATGACTGATGAAAGTGGGAACCCGATTTGGAAAAACCGG 422

Qy        117 ValLysSerCysLysAspLysGluAsnLysLysLysLysArgSerProLysAlaGluThr 136

Db        423 GTCGAAAGTTGGAAAGATAAAAAAGAACAAGAAGAAAAAGGGTGGCGATAAGGTTGCTAAA 482

Qy        137 GluProAlaGlnValProThrGluGlnGlnMetGluGluLysProSerAla--------- 153

Db        483 GAA---GTTCAAGTTCCTGAAGATCAGCATATTGAAGAAAAACAGCAGTCAGCAGACCCA 539

Qy        154 GluAlaSerGluProLeuSerIleValTyrProIleProArgAsnLysLeuThrProTyr 173

Db        540 AATGCAATGCAGCCACCTTCACAAATAATTCCAATTCCAAAGAGCCAAATCACACCTTAT 599

Qy        174 ArgAlaValIleIleMetArgLeuValIleLeuGlyLeuPhePheHisPheArgIleThr 193

Db        600 CGGATTGTGATTATTCTTCGCTTAATCATTTTGGGACTTTTCTTCCATTATAGAATCACA 659

Qy        194 AsnProValAspSerAlaPheGlyLeuTrpLeuThrSerValIleCysGluIleTrpPhe 213

Db        660 AATCCAGTTGAAAGTTCGTACGGTTTATGGCTAACTTCTGTCATTTGTGAAATCTGGTTT 719

Qy        214 AlaPheSerTrpValLeuAspGlnPheProLysTrpAsnProValAsnArgGluThrTyr 233

Db        720 GCTATATCATGGGTTTTGGATCAGTTCCCTAAATGGTATCCCATTAATAGAGTTACATTC 779

Qy        234 IleGluArgLeuSerAlaArgTyrGluArgGluGlyGluProSerGlnLeuAlaGlyVal 253

Db        780 ACAGATGAACTATCTGCAAGGTATGAAAGAGAGGGTGAGCCTTCAGAACTTGCAGCTGTT 839

Qy        254 AspPhePheValSerThrValAspProLeuLysGluProProLeuIleThrAlaAsnThr 273

Db        840 GATTTTTTTGTGTCTACAGTTGATCCACTGAAAGAACCTCCTCTAATTACCGCCAATACC 899

Qy        274 ValLeuSerIleLeuAlaValAspTyrProValAspLysValSerCysTyrValSerAsp 293

Db        900 GTACTTTCAATCCTGGCCGTTGATTACCCCGTTGAAAAAGTCTCCTGTTATGTATCAGAT 959
```

```
 294 AspGlyAlaAlaMetLeuSerPheGluSerLeuValGluThrAlaGluPheAlaArgLys 313
960 GATGGTGCTGCAATGCTCACATTTGAATCTTTAGTTGAAACAGCAGAGTTTGCAAGAAAG 1019

 314 TrpValProPheCysLysLysPheSerIleGluProArgAlaProGluPheTyrPheSer 333
1020 TGGGTCCCATTTTGTAAAAAATTCTCGATAGAACCACGCGCACCTGAGTTTTACTTCTCT 1079

 334 GlnLysIleAspTyrLeuLysAspLysValGlnProSerPheValLysGluArgArgAla 353
1080 CAGAAGATTGACTACTTGAAAGATAAAGTGCAGCCATCGTTTGTCAAAGAACGTAGAGCA 1139

 354 MetLysArgAspTyrGluGluTyrLysValArgValAsnAlaLeuValAlaLysAlaGln 373
1140 ATGAAAAGGGATTACGAAGAGTACAAAGTGCGGGTCAATGCTTTAGTAGCAAAAGCTCAA 1199

 374 LysThrProGluGluGlyTrpThrMetGlnAspGlyThrProTrpProGlyAsnAsnThr 393
1200 AAAACTCCTGAAGAAGGATGGACTATGCAAGATGGAACACCTTGGCCTGGAAATAACCCG 1259

 394 ArgAspHisProGlyHisAspSerGlyLeuProTrpGluIleLeuGlyAlaArgAspIle 413
1260 CGTGACCATCCAGG-CATGATCCAAGTATTTT-TGGGCCATAGT-GGGGCCCATGACATT 1316

 414 GluGlyAsnGluLeuProArgLeuValTyrValSerArgGluLysArgProGlyTyrGln 433
1317 GAAGGAAACGAGCTTCCTCGACTCGTTTATGTGTCCAGAGAAAAGAGGCCAGGATATCAA 1376

 434 HisHisLysLysAlaGlyAlaGluAsnAlaLeuValArgValSerAlaValLeuThrAsn 453
1377 CACCATAAGAAAGCTGGTGCTGAAAATGCTTTGGTGAGGGTGTCTGCGGTTTTGACTAAT 1436

 454 AlaProTyrIleLeuAsnValAspCysAspHisTyrValAsnAsnSerLysAlaValArg 473
1437 GCACCATATATTCTTAATCTGGATTGTGATCACTATGTGAATAATAGTAAGGCGGTTAGA 1496

 474 GluAlaMetCysIleLeuMetAspProGlnValGlyArgAspValCysTyrValGlnPhe 493
1497 GAGGCAATGTGTTTCATGATGGACCCTCAAGTCGGTCGTGATGTGTGCTATATTCAGTTC 1556

 494 ProGlnArgPheAspGlyIleAspLysSerAspArgTyrAlaAsnArgAsnValValPhe 513
1557 CCTCAGAGATTTGATGGCATTGACCGTAGTGACAGATATGCTAATCGGAATACAGTTTTC 1616

 514 PheAspValAsnMetLysGlyLeuAspGlyIleGlnGlyProValTyrValGlyThrGly 533
1617 TTTGATGTTAACATGAAAGGATTAGATGGGATTCAAGGACCGGTTTATGTAGGAACAGGT 1676

 534 CysValPheAsnArgGlnAlaLeuTyrGlyTyrGlyProProSerMetProSerLeuArg 553
1677 TGTGTTTTCTATAGGCAAGCTCTTTACGGTTATGGGCCTCAGTCTTTACCAACATTACCA 1736

 554 LysArgLysAspSerSerSerCysPheSerCysCysCysProSerLysLysLysProAla 573
1737 AGTCCTTCATCGTCGTCGTCTTGT------TGTTGCTGT---GGGCCCAAAAAGCCCAAG 1787

 574 GlnAspProAlaGluValTyrArgAspAlaLysArgGluAspLeuAsnAlaAlaIlePhe 593
1788 AAAGATCTCGAAGAGTTTAAGCGGGATGCCAGAAGGGATGATCTAAACGCTGCCATTTTT 1847

 594 AsnLeuThrGluIleAspAsnTyrAspGluHisGluArgSerMetLeuIleSerGlnLeu 613
1848 AACCTTAAGGAGATAGAAAGTTATGATGATTATGAACGCTCGTTACTCATATCTCAAATG 1907

 614 SerPheGluLysThrPheGlyLeuSerSerValPheIleGluSerThrLeuMetGluAsn 633
1908 AGCTTTGAGAAAACTTTCGGCATGTCTTCTGTGTTTATCGAGTCTACTTTGATGGAAAAT 1967

 634 GlyGlyValProGluSerAlaAsnSerProProPheIleLysGluAlaIleGlnValIle 653
1968 GGAGGATTAGCCGAGTCTGCAAACCCTGCAACTATGATCAATGAAGCAATTCATGTCATT 2027
```

```
Qy   654 GlyCysGlyTyrGluGluLysThrGluTrpGlyLysGlnIleGlyTrpIleTyrGlySer 673
Db  2028 AGTTGTGGATATGAAGAGAAAACCGCATGGGGTAAAGAGATCGGCTGGATTTATGGATCG 2087

Qy   674 ValThrGluAspIleLeuSerGlyPheLysMetHisCysArgGlyTrpArgSerIleTyr 693
Db  2088 GTTACAGAGGATATCTTAACAGGATTCAAGATGCATTGTAGGGGATGGAGATCAATATAC 2147

Qy   694 CysMetProValArgProAlaPheLysGlySerAlaProIleAsnLeuSerAspArgLeu 713
Db  2148 TGCATGCCAGTAAGACCTGCGTTTAAAGGTTCGGCTCCTATCAACTTGTCAGATCGGTTG 2207

Qy   714 HisGlnValLeuArgTrpAlaLeuGlySerValGluIlePhePheSerArgHisCysPro 733
Db  2208 CACCAAGTTTTGAGATGGGCCTTGGATCGGTTGAGATTTTTTTGAGTAGGCATTGCCCG 2267

Qy   734 LeuTrpTyrGlyPheGlyGlyArgLeuLysTrpLeuGlnArgLeuAlaTyrIleAsn 753
Db  2268 TTGTGTATGGTTTTGGAGGAAGGCTTAAATGGCTTCAAAGACTTGCTTATATAAAC 2327

Qy   754 ThrIleValSerVal[illegible]AlaTyrCysThrIleProAlaVal 773
Db  2328 ACCATCGTTTATC[illegible]TGCATACTGCACACCTGCCTGCTATC 2387

Qy   774 CysLeuLeuThrGlyLysPheIleIleProThrLeuSerAsnLeuAlaSerMetLeuPhe 793
Db  2388 TGCCTTCTGACCGGAAAATTCATCATCCCAACGCTTTCCAACATAGCTGCAGTATGGTTT 2447

Qy   794 LeuGlyLeuPheLeuSerIleIleValThrAlaValLeuGluLeuArgTrpSerGlyVal 813
Db  2448 CTTGGGCTCTTTTTGTCAATTATTACAACCAGTGTTCTTGAAATTCGATGGAGCGGTGTG 2507

Qy   814 SerIleGluAspLeuTrpArgAsnGluGlnPheTrpValIleGlyGlyValSerAlaHis 833
Db  2508 AGCATCGAAGAGCTTTGGAGAAACGAGCAATTTTGGGTCATCGGTGGTGTCTCCGCTCAT 2567

Qy   834 LeuPheAlaValPheGlnGlyPheLeuLysMetLeuAlaGlyIleAspThrAsnPheThr 853
Db  2568 CTGTTTGCTGTTTTTCAAGGCTTCCTGAAAATGCTAGCTGGTGTGGATACCAACTTCACT 2627

Qy   854 ValThrAlaLysAlaAlaGluAspAlaGluPheGlyGluLeuTyrMetValLysTrpThr 873
Db  2628 GTCACTGCAAAAGCAGCTGATGATCAGGAATTTGGGGAGCTTTATATGATCAAATGGACA 2687

Qy   874 ThrLeuLeuIlePr[illegible]ProThrThrLeuLeuIleIleAsnMetSerGly-  CysAlaGly 892
Db  2688 ACGGTTTTAATCCGTCCGACCACCCTGCTGGTACTGAACTTGGTAGGCGTTGTTGCAGGG 2747

Qy   893 PheSerAspAlaLeuAsnLysGlyTyrGluAlaTrpGlyProLeuPheGlyLysValPhe 912
Db  2748 TTTTCGGATGCGTTAAACAAAGGGTATGAGGCTTGGGGCCCGTTGTTTGGGAAAGTGTTC 2807

Qy   913 PheAlaPheTrpValIleLeuHisLeuTyrProPheLeuLysGlyLeuMetGlyArgGln 932
Db  2808 TTTGCGTTCTGGGTCATTCTTCATTTGTACCCGTTTCTTAAAGGTCTCATGGGTCGCCAG 2867

Qy   933 AsnLeuThrProThrIleValValLeuTrpSerValLeuLeuAlaSerValPheSerLeu 952
Db  2868 AACCGGACTCCCACCATTGTTATCTTGTGGTCGGTGTTGTTGGCGTCTGTTTTCTCACTT 2927

Qy   953 ValTrpValLysIleAsnProPheValAsnLysValAspAsnThrLeuValAlaGluThr 972
Db  2928 GTGTGGGTGAAGATTGATCCGTTCGTTAGCAAAGGTGATTCCAATCTG---ACACAAGGA 2984

Qy   973 CysIleSerIleAspCys 978
Db  2985 TGCATTGCAATTGATTGT 3002

RESULT 7
AF413210
LOCUS       AF413210               6886 bp    DNA     linear   PLN 27-DEC-2001
DEFINITION  Gossypium hirsutum cellulose synthase A4 (CesA4) gene, complete
            cds.
```

```
ACCESSION   AF413210
VERSION     AF413210.1  GI:17226293
KEYWORDS    .
SOURCE      Gossypium hirsutum (upland cotton)
  ORGANISM  Gossypium hirsutum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium.
REFERENCE   1  (bases 1 to 6886)
  AUTHORS   Kim,H.J. and Triplett,B.A.
  TITLE     Cotton fiber growth in planta and in vitro. Models for plant cell
            elongation and cell wall biogenesis
  JOURNAL   Plant Physiol. 127 (4), 1361-1366 (2001)
  MEDLINE   21608363
   PUBMED   11743074
REFERENCE   2  (bases 1 to 6886)
  AUTHORS   Kim,H.J., Williams,M.Y. and Triplett,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-AUG-2001) Fiber Bioscience, USDA-ARS, SRRC, 1100
            Robert E. Lee Blvd, New Orleans, LA 70124, USA
FEATURES             Location/Qualifiers
     source          1. .6886
                     /organism="Gossypium hirsutum"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3635"
     gene            <2685. .>6766
                     /gene="CesA4"
     mRNA            join(<2685. .2862,3033. .3106,3193. .3244,3338. .3482,
                     3613. .3879,3992. .4337,4426. .4689,4778. .4990,5085. .5328,
                     5420. .5619,5732 .6085,6179. .>6766)
                     /gene="CesA4"
                     /product="cellulose synthase A4"
     CDS             join(2685. .2862,3033. .3106,3193. .3244,3338. .3482,
                     3613. .3879,3992. .4337,4426. .4689,4778. .4990,5085. .5328,
                     5420. .5619,5732 .6085,6179. .6766)
                     /gene="CesA4"
                     /codon_start=1
                     /product="cellulose synthase A4"
                     /protein_id="AAL37718.1"
                     /db_xref="GI:17226294"
                     /translation="MMESGVPVCHTCGEHVGLNVNGEPFVACHECNFPICKSCFEYDL
                     KEGRKACLRCGSPYDENLLDDVEKTTGDQSTMAAHLSKSQDVGIHARHISSVSTLDSE
                     MTGDNGNPIWKNRVESWKEKKNKKKKPATTKVERKAEIPPEQQMEDKPAPDASQPLST
                     IIPIPKSRLAPYRTVIIMRLIILGLFFHYRVTHPVDSAFGLWLTSVICEIWFAFSWVL
                     DQFPKWYPVNRETYIDRLSARYDREGEPNELAAVDFFVSTVDPLKEPPLITANTVLSI
                     LALDYPVDKVSCYISDDGAAMLTFESLVETADFARKWVPFCKKFSIEPRAPEFYFSQK
                     IVYLKDKVQPSFVKERRAMKREYEEYKIRINALVAKAQKTPEEGWTMQDGTPWPGNNP
                     RDHPGMIQVFLGYSGAHDTEGNELPRLVYVSREKRPGYQHHKQAGADNALVRVSAALT
                     NAPFILNLDCDHYVNNSKAVREAMCFLMDQGDEICYVQFPQRFDGIDRSDRYANRN
                     TVFFDVNMKGRDGSQGIYYVGTGCVCNREALYGYDPPSVSEPKSSSSSCSCCCPGKK
                     EPKEPTELYRDAKREELDAAIFKLREIDNYDEYKDDMLISQTSFEKTFGLSSVFIEST
                     LMENGGVAESANPSTLIKEAIHVIGCGYEEKTAWGKEIGWIYGSVTEDILTGFKMHCR
                     GWRSIYCMPLRPAFKGSAPINLSDRLHQVLRWALGSLEIFLSRHCPLWYGFGGGRLKW
                     LQRLAYINTSVYPFTSLPLIAYCSLPAICLLTGKFIIPTLSNLASVLYLGLFLSIIVT
                     AVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQGFLKMLAGIDTNFTVTAKAADD
                     ADFGELYIVKWTTLLIPPTTLLIVNMVGVVAGFSDAINKGYEAWGPLFGKVFFSFWVI
                     LHLYPFLKGLMGRQNRTPTIVVLWSVLLASVFSLVWVRINPFVSTADSTTVSRANSSI
                     DC"
BASE COUNT     2067 a    1237 c    1277 g    2304 t       1 others
ORIGIN

Alignment Scores:
Pred. No.:              0             Length:         6886
Score:                  3990.00       Matches:        626
Percent Similarity:     65.65%        Conservative:   74
Best Local Similarity:  60.25%        Mismatches:     73
Query Match:            76.15%        Indels:         42
DB:                     8             Gaps:           17

US-09-980-043-2 (1-978) x AF413210 (1-6886)

Qy          1 MetMetGluSerGlyAlaProIleCysHisThrCysGlyGluGlnValGlyHisAspAla 20
              ||||||||||||||||  | |:::|||||.|||||:|||  :  :  |||   :::

Db       2685 ATGATGGAATCTGGGGTTCCTGTTTGCCACACTTGTGGTGAACATGTTGGGTTGAATGTT 2744

Qy         21 AsnGlyGluLeuPheValAlaCysHisGluCysSerTyrProMetCysLysSerCysPhe 40
              |||||||||  |||||||||||||||||||||:::::|||:::|||||||||||||||

Db       2745 AATGGTGAACCCTTTGTGGCTTGCCATGAATGTAATTTCCCTATTTGTAAGAGTTGTTTT 2804

Qy         41 GluPheGluIleAsnGluGlyArgLysValCysLeuArgCysGlySerProTyrAspGlu 60
              |||:::::::::   |||| |||||||  |||||||||||||||||||||||||||||

Db       2805 GAGTATGATCTTAAGGAAGGACGAAAAGCTTGCTTGCGTTGTGGTAGTCCATATGATGG- 2863

Qy         60 ------------------------------------------------------------ 60

Db       2864 TACATTCATTTTAAAATCGTGATGCTAGTGTTTTACGTGTTTCCATGTGGTTCTTGATAA 2923

Qy         60 ------------------------------------------------------------ 60

Db       2924 ATCTCTGTTATAATCTTGTGTTTGCCAAGTGCTTGTTTGATTTCCCATGAGAAAGATAGA 2983

Qy         61 ----------------------------------------------------AsnLeuLeu 63
                                                                  ||||||||

Db       2984 GATAGAAGTTGTAAACTGATTGCACGACTGATTATTGTGAACTCCTCAGAAAACCTGTTG 3043

Qy         64 AspAspValGluLysLysGlySerGlyAsnGlnSerThrMetAlaSerHisLeuAsnAsp 83
              |||||||||||||||      :::|||:::|||||||||||||||:::|||||||:::

Db       3044 GACGATGTCGAGAAGACC---ACCGGCGATCAATCGACAATGGCTGCACATTTGAGCAAG 3100

Qy         84 Ser--------------------------------------------------------- 84

Db       3101 TC-TCAGGTATTGTTTTTTGCTGCTTTTGAAAGGTTCAGTCAAAGCCAGGAGGGATATTG 3159

Qy         85 ----------------------------GlnAspValGlyIleHisAlaArgHisIle 94
                                          |||||||||||||||||||||||||||||

Db       3160 AGAGCTTTGTGTTATTCATGTTGGTTTATACAGGATGTTGGAATTCATGCAAGACATATC 3219

Qy         95 SerSerValSerThrValAspSer---------------------------------- 102
              ||||||||||||||:::|||||

Db       3220 AGCAGTGTGTCTACATTGGATAGTGGTATGTAATATAAATTTCTTTCCATATTATTGGTT 3279

Qy        103 ----------------------------------------------------------Glu 103
                                                                        |||

Db       3280 TCTTTCTTTCAATTTCCTTCAGTTCTGAAAGCTCGAATCGTCTCGTAAAATATCGTAGAA 3339

Qy        104 MetAsnAspGluTyrGlyAsnProIleTrpLysAsnArgValLysSerCysLysAspLys 123
              |||     :::  ||||||||||||||||||||||||||:::|||  |||:::|||

Db       3340 ATGACTGGAGACAATGGGAATCCGATTTGGAAGAACAGGGTGGAAAGTTGGAAAGAAAAG 3399

Qy        124 GluAsnLysLysLysLysArgSerProLysAlaGluThrGluProAlaGlnValProThr 143
              :::|||||||||||||||  :::                 |||:::::|||

Db       3400 AAGAACAAGAAGAAGAAGCCTGCAACAACTAAGGTTGAAAGAGAGGCTGAAATCCCACCT 3459

Qy        144 GluGlnGlnMetGluGluLysPro----------------------------------- 151
              |||||||||||||||:::||||||

Db       3460 GAGCAACAAATGGAAGATAAACCGTAAGTATTTAATTAGAAAAGCCTTGCTTGATTTTGA 3519

Qy        151 ------------------------------------------------------------ 151

Db       3520 AATATACAAAGCAGATTAATGATAAGAGCCTATTTAGTGAAATGATTTTTGTTTATGTTA 3579

Qy        152 ---------------------------------SerAlaGluAlaSerGluProLeuSe 160
                                                :::  :::||||||:::|||||||

Db       3580 TGTGTTACTGATATGTAATATTTGCATCTGCAGGGCACCGGATGCTTCCCAGCCCCTCTC 3639

Qy        160 rIleValTyrProIleProArgAsnLysLeuThrProTyrArgAlaValIleIleMetAr 180
              |  :::  |||||||||:::::::::|||  |||||||||  ||||||||||||||

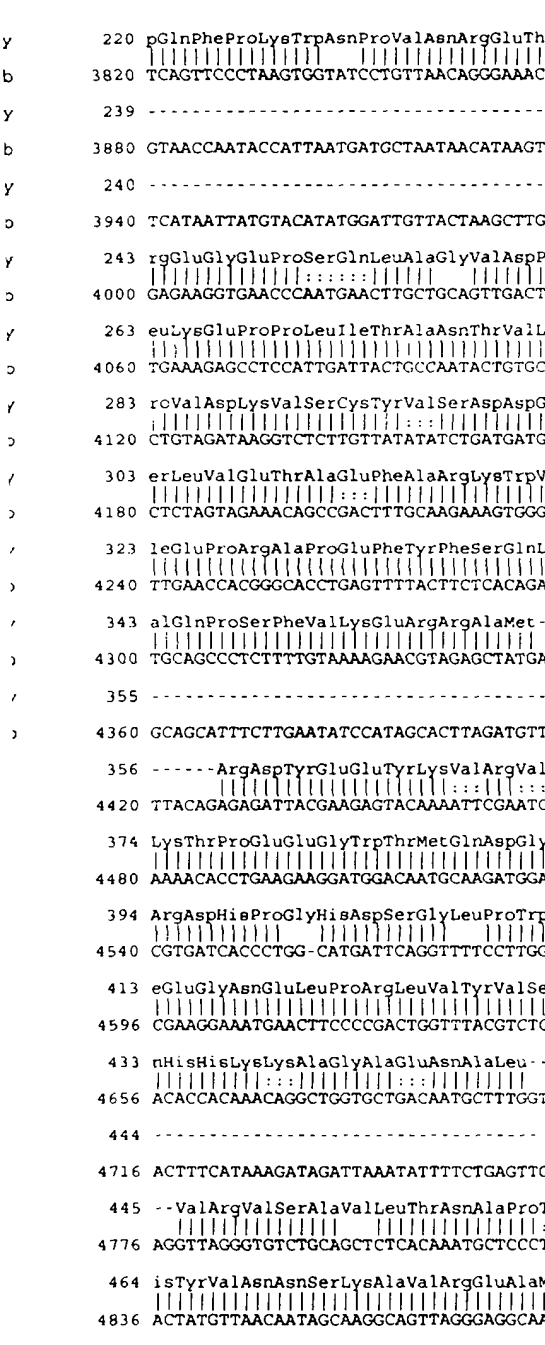
Db       3640 GACTATAATTCCAATCCCGAAAAGCAGACTTGCACCATACCGAACCGTGATCATTATGCG 3699

Qy        180 gLeuValIleLeuGlyLeuPhePheHisPheArgIleThrAsnProValAspSerAlaPh 200
              ||||:::|||||||||||||||'||||||:::|||:::|||:::||||||||||||||||

Db       3700 ATTGATCATTCTCGGTCTTTTCTTCCATTATCGAGTAACACACCCCGTTGACAGTGCTTT 3759

Qy        200 eGlyLeuTrpLeuThrSerValIleCysGluIleTrpPheAlaPheSerTrpValLeuAs 220
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db       3760 TGGACTGTGGCTCACTTCAGTCATATGTGAAATCTGGTTTGCTTTCTCCTGGGTGTTGGA 3819
```





```
                    ['|||||||||||||||||||||||||||||||    ||||||||||||||||||||||||||||||
Db       5967 GGCTTCAAAGACTAGCATATATAAACACCAGTGTCTATCCTTTCACATCCCTCCCACTCA 6026

Qy        765 leAlaTyrCysThrIleProAlaValCysLeuLeuThrGlyLysPheIleIleProThr- 784
              |||||||||||:::::|||||||:|:|||||||||||||||||||||||||||||||||
Db       6027 TTGCCTATTGTTCACTACCAGCAATCTGTCTTCTCACAGGAAAATTTATCATACCAACGG 6086

Qy        784 ---- --  ------------  -- --  -- --- --  --- --  --------- 784

Db       6087 TAATGAATGCTAGTAATACATGACTTCTTTTTTGAGTTTTCAATACCATATATATGCTAC 6146

Qy        785 --- --- ---------- --- --- -- -LeuSerAsnLeuAlaSerMetLeuPheL 794
                                            |:|||||||||||||||:::|||:::|
Db       6147 CACTGCTTCACATTTTCTAATGTGATGATCAGCTCTCAAACCTGGCAAGTGTTCTCTATC 6206

Qy        794 euGlyLeuPheIleSerIleIleValThrAlaValLeuGluLeuArgTrpSerGlyValS 814
              |||||||||:::|||||||||||:|||||||||||||||||||||||||||||||||||
Db       6207 TTGGCCTTTTCCTTTCCATTATCGTGACCGCTGTTCTCGAGCTCCGATGGAGTGGTGTCA 6266

Qy        814 erIleGluAspLeuTrpArgAsnGluGlnPheTrpValIleGlyGlyValSerAlaHisL 834
              |||||||||||||||||||||:||||||:|||||||||||||||||||||||||||||||
Db       6267 GCATCGAGGACTTATGGCGTAACGAGCAGTTTGGGTCATCGGTGGCGTTTCAGCGCATC 6326

Qy        834 euPheAlaValPheGlnGlyPheLeuLysMetLeuAlaGlyIleAspThrAsnPheThrV 854
              |||||||||||||||||:|||||||||||||||||||||||||||||||||||||||||
Db       6327 TCTTTGCCGTCTTCCAAGGTTTCCTTAAGATGCTTGCAGGCATTGACACCAACTTTACTG 6386

Qy        854 alThrAlaLysAlaAlaGluAspAlaGluPheGlyGluLeuTyrMetValLysTrpThrT 874
              |||||||||||||||||:::|||.||:::|||||||||||:::|||||||||||||||||
Db       6387 TCACTGCCAAAGCAGCTGATGATGCAGATTTTGGTGAGCTCTACATTGTGAAATGGACTA 6446

Qy        874 hrLeuLeuIleProProThrThrLeuLeuIleIleAsnMetSerGly---CysAlaGlyP 893
              |:|||||||||||||||||||||||| ||||:::|||||   ||:       |||||||
Db       6447 CACTTTTAATCCCTCCAACAACACTCCTCATCGTCAACATGGTTGGTGTCGTTGCCGGAT 6506

Qy        893 heSerAspAlaLeuAsnLysGlyTyrGluAlaTrpGlyProLeuPheGlyLysValPheP 913
              |||||||:|||||||||||||||||:||||:|||:|||||||||||||||||||||||||
Db       6507 TCTCTGATGCCCTCAACAAAGGGTACGAAGCTTGGGGACCACTCTTTGGCAAAGTGTTCT 6566

Qy        913 heAlaPheTrpValIleLeuHisLeuTyrProPheLeuLysGlyLeuMetGlyArgGlnA 933
              ||:::|||||||||||||||||||| |||:|||||:|||:||| ||||||||||||||||
Db       6567 TTTCCTTCTGGGTCATCCTCCATCTTTATCCATTCCTCAAAGGTCTTATGGGACGGCAAA 6626

Qy        933 snLeuThrProThrIleValVal--TrpSerValLeuLeuAlaSer-ValPheSerLeuV 953
              |:  |||:|||||||||||||  ||| :|||  |||. |:| :|:| ||||||||||||
Db       6627 ACAGGACACCAACCATTGTTGTC TTCCTTTACTCTTGTCTTCTC-TTTTCTCTTG 6686

Qy        953 alTrpValLysIleAsnProPheValAsnLysValAspAsnThrLeuValAlaGluThrC 973
              ||||||| :::|||||||||:|||||:::      |'|:::  |   :|:|:::
Db       6687 TTTGGGTTCGGATCAACCCGTTTGTCAGCAAGGCGATAGCAACACTGTGTCAAGGGCGA 6746

Qy        973 ysIleSerIleAspCys 978
                ||||||||||||
Db       6747 ATTCTTCCATTGATTGT 6763

RESULT 8
ATF28A21/c
LOCUS       ATF28A21               94301 bp    DNA     linear   PLN 08-SEP-1999
DEFINITION  Arabidopsis thaliana DNA chromosome 4, BAC clone F28A21 (ESSA
            project).
ACCESSION   AL035526
VERSION     AL035526.1  GI:4539378
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Bevan,M., Mueller,M.W., Muendlein,A., Felber,R., Bancroft,I.,
            Mewes,H.W., Mayer,K.F.X., Lemcke,K. and Schueller,C.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 94301)
  AUTHORS   EU Arabidopsis sequencing,project.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-SEP-1999) MIPS, at the Max-Planck-Institut fuer
            Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
            schuelle@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de Project
            Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
            Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
            E-mail: michael.bevan@bbsrc.ac.uk
COMMENT     Information on performance of analysis and a more detailed
            annotation of this entry and other sequences of chromosomes 3, 4
            and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/.
FEATURES             Location/Qualifiers
     source          1. .94301
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /variety="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="4"
     gene            9652. .10227
                     /gene="F28A21.20"
     CDS             9652. .10227
                     /gene="F28A21.20"
                     /note="similarity to predicted proteins, Arabidopsis
                     thaliana
                     contains EST gb:T46643, R29928"
                     /codon_start=1
                     /product="putative protein"
                     /protein_id="CAB37446.1"
                     /db_xref="GI:4539380"
                     /translation="MSSDRHTPTKDPPDHPSSSSNHHKQPLPPQPQQPLSRYESQKRR
                     DWNTFVQYLKSQNPPLMMSQFDYTHVLSFLRYLDQFGKTKVHHQACVFFGQPDPPGPC
                     TCPLKQAWGSLDALIGRLRAAYEEHGGGSPDTNPFANGSIRVHLREVRESQAKARGIP
                     YRKKKRRKTKNEVVVVKKDVANSSTPNQSFT"
     exon            9652. .10227
                     /gene="F28A21.20"
                     /number=1
     gene            13217. .13711
                     /gene="F28A21.30"
     CDS             13217. .13711
                     /gene="F28A21.30"
                     /note="similarity to various predicted proteins,
                     Arabidopsis thaliana"
                     /codon_start=1
                     /product="putative protein"
                     /protein_id="CAB37447.1"
                     /db_xref="GI:4539381"
                     /translation="MESSKQKRCRSSVVETIEAPLPLVWSILRSFDKPQAYQRFVKSC
                     TMRSGGGGGKGGEGKGSVRDVTLVSGFPADFSTERLEELDDESHVMVVSIIGGNHRLV
                     NYKSKTKVVASPEDMAKKTVVVESYVVDVPEGTSEEDTIFFVDNIIRYNLTSLAKLTK
                     KMMK"
     exon            13217. .13711
                     /gene="F28A21.30"
                     /number=1
     gene            16036. .17454
                     /gene="F28A21.40"
     CDS             16036. .17454
                     /gene="F28A21.40"
                     /note="similarity to predicted protein, Arabidopsis
                     thaliana"
                     /codon_start=1
                     /product="putative protein"
                     /protein_id="CAB37448.1"
                     /db_xref="GI:4539382"
                     /translation="MEGKKLNLYAPLPSIRRIPSMIERSSELENKKTTITRPELRSCP
                     STKQETPVFVLPDQSFDHLTEPASIPFMWEQIPGKPKDDMATLIQESGLLETDDEEED
                     EVDEDLDTVSSNVSFSVNCSTSGVSEIEKTGERSDCDEKSRESLDLMMSRFLPAAKAM
                     ALQTHQKHQSSYNSSEQKLITQNREALVARQRSQLVAEHEHFAIVQSLYDDLNIDDDT
                     EDDENDDDGDDYGHVDHKIYPAEVTKKACGFLPRLCAKNSFKFLNPVPLDSVSNRSMK
                     HSGDQTSLPNWSTRRLSGFISPYRTSCSDSGFLGTPEKPESFKRLNRGISKSQELYPT
                     RTRREVLPNYSNSGEIKMFRNSISTPSRIQRTTIHDSRFLVEEVNRRRNNKSGNLQQT
                     SSEYTPALSPPPLPETPSRSWLGRTLLPPVNPRPYGVVLGQVGIKKLNQEVLESTKWE
                     TIVKTSYVHNDHARYSQVLITL"
     exon            16036. .17454
```

```
                     /gene="F28A21.40"
                     /number=1
     gene            19501. .22435
                     /gene="F28A21.50"
     CDS             join(<19501. .19652,19742. .19813,19902. .20045,20144. .20287,
                     20375. .20469,20559. .20987,21071. .21261,21342. 21474,
                     21531. .21709,21893. .22084,22178. .22435)
                     /gene="F28A21.50"
                     /codon_start=1
                     /product="putative protein (fragment)"
                     /protein_id="CAB37449.1"
                     /db_xref="GI:4539383"
                     /translation="FVLNSGSALLKFRARVNSDPHGTLANWNVSGINDLCYWSGVTCV
                     DGKVQILDLSGYSLEGTLAPELSQLSDLRSLILSRNHFSGGIPKEYGSFENLEVLDLR
                     ENDLSGQIPPELSNGLSLKHLLLSGNKFSDDMRIKIVRLQSSYEVRLKKSPKLSPLAV
                     LGCINRKLGHCVSRNRIIQVKKVEAIVFRIKATSRRFLKAPPSFLEETDIYKRRELLE
                     ETSNLAAEPAPSAPSPSPGIITEASPRSSGSFPAVTNAKKRRPPLVPPVPSPDKGSTS
                     PDISKNQPQDNKQSKGSKHVWLYVVIAVASFVGLLIIVAVIFFCRKRAVKSIGPWKTG
                     LSGQLQKAFVTGVPKLNRSELETACEDFSNIIETFDGYTVYKGTLSSGVEIAVASTAI
                     AESKEWTRAMEMAYRRKIDTLSRINHKNFVNLIGYCEEDDPFNRMMVFEYAPNGTLFE
                     HLHVNNVFRLSDKETEHLDWSARMRIIMGTAYCLQHMHGMNPPMAHTDFNSSEIYLTD
                     DYAAKVSEIPFNLEARLNPKKHVSGDLEQTSLLLPPEPEANVHSFGVLMLEIISGKLS
                     FSDEYGSIEQWASKYLEKDDLGEMIDPSLKTFKEEELEVICDVIRECLFTEQPLRPSM
                     KDVAEQLKQVINITPEKATPRSSPLWWAELEILSSEAT"
     exon            19501. .19652
                     /gene="F28A21.50"
                     /number=1
     intron          19653. .19741
                     /gene="F28A21.50"
                     /number=1
     exon            19742. .19813
                     /gene="F28A21.50"
                     /number=2
     intron          19814. .19901
                     /gene="F28A21.50"
                     /number=2
     exon            19902. .20045
                     /gene="F28A21.50"
                     /number=3
     intron          20046. .20143
                     /gene="F28A21.50"
                     /number=3
     exon            20144. .20287
                     /gene="F28A21.50"
                     /number=4
     intron          20288. .20374
                     /gene="F28A21.50"
                     /number=4
     exon            20375. .20469
                     /gene="F28A21.50"
                     /number=5
     intron          20470. .20558
                     /gene="F28A21.50"
                     /number=5
     exon            20559. .20987
                     /gene="F28A21.50"
                     /number=6
     intron          20988. .21070
                     /gene="F28A21.50"
                     /number=6
     exon            21071. .21261
                     /gene="F28A21.50"
                     /number=7
     intron          21262. .21341
                     /gene="F28A21.50"
                     /number=7
     exon            21342. .21474
                     /gene="F28A21.50"
                     /number=8
     intron          21475. .21530
                     /gene="F28A21.50"
                     /number=8
     exon            21531. .21709
                     /gene="F28A21.50"
                     /number=9
     intron          21710. .21892
                     /gene="F28A21.50"
                     /number=9
     exon            21893. .22084
                     /gene="F28A21.50"
                     /number=10
     intron          22085. .22177
                     /gene="F28A21.50"
                     /number=10
     exon            22178. .22435
                     /gene="F28A21.50"
                     /number=11
     gene            22901. .24271
                     /gene="F28A21.60"
     gene            complement(join(22901. .23176,23858. .24271))
                     /gene="F28A21.60"
     CDS             complement(join(22901. .23176,23858. .24271))
                     /gene="F28A21.60"
                     /note="similarity to transcription factor OBF1.1, Zea
                     mays, EMBL:AJ238534"
                     /codon_start=1
                     /product="putative protein"
                     /protein_id="CAB37450.1"
                     /db_xref="GI:4539384"
                     /translation="MRNDVEEKFLEFYESWVIQLELYLHQLLIAHNNNTMSETELRHL
                     ISKLTTHHKAYYTAKWAARPEDVLAFPGSVWLNPLENACSWLTGWKPSMVFRMVDRLR
                     KSRVVLVFAQVKKLEEIRVETKPDEQKIEREMERYQVAMADRKVVELARIGCHVGGES
                     VMVVEAARGLSMGLFKMVEAADCVRGKTLKGILDILTPPQCVEFLAAAATFQVQLRR
                     WGNRRHFVTHS"
     exon            complement(22901. .23176)
                     /gene="F28A21.60"
                     /number=1
     intron          complement(23177. .23857)
                     /gene="F28A21.60"
                     /number=1
     exon            complement(23858. .24271)
```

```
Alignment Scores:
Pred. No.:                0                 Length:          94301
Score:                    3926.00           Matches:           806
Percent Similarity:       68.6%             Conservative:       87
Best Local Similarity:    62.0%             Mismatches:         81
Query Match:              74.9%             Indels:            331
DB:                       8                 Gaps:               16

US-09-980-043 2 (1-978) x AF028A21 (1-94301)

Qy          1 MetMetGluSerGlyAlaProIleCysHisThrCysGlyGluGlnValGlyHisAspAla 20
              |||  |||  ||   :::|||||||| |::::.|||||  |||||::::::|:|:      :::
Db      75577 ATGATGGAGTCTAGGTCTCCCATCTGCAACACTTGTGGTGAAGAGATTGGTGTAAAATCA 75518

Qy         21 AsnGlyGluLeuPheValAlaCysHisGluCysSerTyrProMetCysLysSerCysPhe 40
              :|  |||   :  ||||:|||. |||:|:|:| :::||||::::|| .|||:::.||
Db      75517 AACGGAGAGTTCTTTGTGGCTTGTCATGAGTGTAGTTTCCCGATCTGCAAAAGCTTGTCTT 75458

Qy         41 GluPheGluIleAspGluGlyArgLysValCysLeuArgCysGlySerProTyrAspGlu 60
              |:::::.:      |:|||.:::::::||   ||||:||||::|||:.|||| 
Db      75457 GAGTATGAATTCAAAGAAGGTCGAAGAACTTGCCTGCGTTGCGGCAATCCTTACGATGG  75399

Qy         60 ------   ---  :  ----      :      ------------  -----   -- 60

Db      75398 TACAATTATAAATTTACAAACTTTATCGAAAGTTTGATTTTTTGGTTAAAGAGTATTTA 75339

Qy         61 --   --------  ------  - -   --- AsnLeuLeuAspAspValGluLysLys 69
                                                |||:::  |||||||:|||  |||
Db      75338 TAACTTGTTTTGTTTTGTTTTGCTTTCATAGAGAATGTGTTTGATGATGTTGAGACAAAG 75279

Qy         70 GlySerGlyAsnGlnSerThrMetAlaSerHisLeuAsnAspSer-- --------  - 84
                |||      ||||||  :::  :::      !||:::
Db      75278 ACATCTAAAACTCAATCCATTGTTCGAACACAGACCAATAACAC-TTCTCAGGTTTGTGT 75220
```



```
Qy        903 AlaTrpGlyProLe PheGlyLysVa lleProPheAlaPheTrpValIleLeuHisLeuTyr 922
              ||||||||||:|||   ||  |||  | |  :||.  ,|||||:|||||,:||||,|||||
Db      71934 GCTTGGGGACCTTC TTTGGGAAGTCATTTT GCCTTTTGGGTGATTCTTCATCTTTAT 71875

Qy        923 ProPheLeuLysGl LeuMetGlyArgGlnAs LeuThrProThrIleValValLeuTrp 942
              |||||||  |||: | |  ||  :||. |||     ||||||||||.||||:::||:: |
Db      71874 CCATTCCTCAAAGG TTATGGGAAAATAAAA AGAACACCTACTATTGTTATTCTCTGG 71815

Qy        943 SerValLeuLeuAl erValPheGl ValTrpValLysIleAsnProPheValAsn 962
              ||:::  |||  ,|:  |||  :||||. ||  ||||||:::||||||||||||:::
Db      71814 TCTATCTTGCTTGC TGTGTTTT ACTTGTTGGGTTCGTATCAATCCTTTCGTCTCC 71755

Qy        963 LysValArgAsnTy LeuValAlaGluThrCysIleSerIleAspCys 978
              ||  |||   ||  :::::::  '  :::   ||||||:|||
Db      71754 AAAACGGATACGA  TTTTTTT  AAACTCTTTTGATCGATTGC 71707

RESULT 9
ATCHRIV49/c
LOCUS       ATCHRIV49              linear   PLN 16-MAR 2000
DEFINITION  Arabi                  , working fragment No. 49.
ACCESSION   AL161549
VERSION     AL161549.2  GI:
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Brassicales; Brassicaceae; Arabidopsis
REFERENCE   1  (bases 1 to )
  AUTHORS   Hilbert,H., Brau ,M., Holper,F. Brandt,A., Duesterhoeft,A.,
            Mewes,H.W., Lemcke,K. and Mayer,K.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to )
  AUTHORS   Mueller,M.W., Mundlein,A., Pohl,T., Mewes,H.W., Lemcke,K. and
            Mayer,K.F.X.
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 118148)
  AUTHORS   Robben,J., Grymp ez,B., Volckaert,G., Mewes,H.W., Lemcke,K. and
            Mayer,K.F.X.
  JOURNAL   Unpublished
REFERENCE   4  (bases 1 to 199075)
  AUTHORS   Pohl,T., Weizenegger,T., Mewes,H.W., Lemcke,K. and Mayer,K.F.X.
  JOURNAL   Unpublished
REFERENCE   5  (bases 1 to 199075)
  AUTHORS   EU Arabidopsis sequencing project.
  TITLE     Direct Submission
  JOURNAL   Submitted (10 MAR 2000) MIPS, at the Max-Planck-Institut fuer
            Biochemie, Am Klopferspitz 18a, D 82152 Martinsried, FRG, E-mail:
            lemcke@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de Project
            Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
            Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
            E-mail: michael.bevan@bbsrc.ac.uk
COMMENT     Information on performance of analysis and a more detailed
            annotation of this entry and other sequences of chromosomes 3, 4
            and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/
            this fragment has an overlap with ATCHRIV48 at the 5' end and an
            overlap with ATCHRIV50 at the 3' end.
FEATURES             Location/Qualifiers
     source          1. .199075
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /variety="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="4"
     gene            9577. .12672
                     /gene="AT4g18560"
     CDS             join(9577. .9906,9958. .10809,10980. .11111,11160. .11377,
                     11463. .11574,12310. .12447,12541. .12672)
                     /gene="AT4g18560"
                     /note="similarity to pherophorin-S, Volvox carteri,
                     PIR:T10798
                     contains EST gb:AI997040.1, W43771"

73073 CTTCCGAGATTTACAAAGATGCAAAGCGAGAAGAACTTGATGCTGCAATCTTTAATCTTG 73014
  596 hrGluIleAspAsnTyrAspGluHisGluArgSerMetLeuIleSerGlnLeuSerPheG 616
73013 GGGACCTAGACAACTACGATGAGTACGACAGATCGATGCTGATTTCACAAACAAGCTTTG 72954
  616 luLysThrPheGlyLeuSerSerValPheIleGluSerThrLeuMetGluAsnGlyGlyV 636
72953 AGAAAACGTTTGGTCTCTCTACGGTTTTCATCGAGTCTACTCTTATGGAGAATGGCGGTG 72894
  636 alProGluSerAlaAsnSerProProPheIleLysGluAlaIleGlnValIleGlyCysG 656
72893 TTCCCGACTCTGTAAACCCGTCAACGCTCATCAAAGAAGCTATTCATGTCATTAGCTGTG 72834
  656 lyTyrGluGluLysThrGluTrpGly-Lys------- --------- --------  - 665
72833 GATACGAAGAGAAAACTGAATGGGGAAAAGAAGTAAGTGCCTTGTTTCTTCATATCAAAG 72774
  666 --- ---------- -- --- - ---GlnIleGlyTrpIleLys 671
72773 TTCTCTTGTGAGTACATTTTATAACCTTTCGTTCTGTTTTCGCAGATAGGATGGATTTAC 72714
  672 GlySerValThrGluAspIleLeuSerGlyPheLysMetHisCysArgGlyTrpArgSer 691
72713 GGGTCGATCACCGAAGACATTTTGACGGGTTTCAAGATGCATTGTCGTGGATGGAGGTCG 72654
  692 IleTyrCysMetProValArgProAlaPheLysGlySerAlaProIleAsnLeuSerAsp 711
72653 ATTTACTGTATGCCATTAAGACCAGCATTTAAAGGATCTGCTCCAATCAATCTATCAGAT 72594
  712 ArgLeuHisGlnValLeuArgTrpAlaLeuGlySerValGluIlePhePheSerArgHis 731
72593 AGGCTTCACCAGGTTCTACGTTGGGCTCTTGGCTCGGTTGAGATCTTCCTTAGCCGACAT 72534
  732 CysProLeuTrpTyrGlyPheGlyGlyGlyArgLeuLysTrpLeuGlnArgLeuAlaTyr 751
72533 TGTCCTTTGTGGTACGGTTGCAGCGGAGGCCGCCTCAAGTTGCTCCAGAGATTAGCTTAT 72474
  752 IleAsnThrIleValTyrProPheThrSerLeuProLeuIleAlaTyrCysThrIlePro 771
72473 ATAAACACTATTGTCTACCCATTCACTTCTTTGCCTCTTGTTGCTTACTGTACTCTTCCA 72414
  772 AlaValCysLeuLeuThrGlyLysPheIleIleProThrLeu------------  --- 785
72413 GCTATTTGCCTTCTTACCGGCAAATTTATCATCCCAACGGT-AACTTTTTCATACTACTA 72355
  785 ---------------------------------- ------------------- 785
72354 ATTACTTTGGTTAGGTTACAGCAAGATTGAGAATTTGATCTTTGTTTCTGTTCTTTATTG 72295
  786  -----SerAsnLeuAlaSerMetLeuPheLeuGlyLeuPheIleSerIleIleValThr 803
72294 CAGCTATCAAACCTAGCAAGCATGCTGTTTCTAGGTCTCTTTATATCAATCATCTTAACG 72235
  804 AlaValLeuGluLeuArgTrpSerGlyValSerIleGluAspLeuTrpArgAsnGluGln 823
72234 AGTGTCCTCGAGCTTCGATGGAGCGGAGTCAGTATCGAAGACTTATGGAGAAACGAACAG 72175
  824 PheTrpValIleGlyGlyValSerAlaHisLeuPheAlaValPheGlnGlyPheLeuLys 843
72174 TTTTGGGTTATTGGAGGTGTCTCAGCTCATCTCTTTGCCGTTTTCCAAGGATTCCTCAAA 72115
  844 MetLeuAlaGlyIleAspThrAsnPheThrValThrAlaLysAlaAlaGluAspAlaGlu 863
72114 ATGCTCGCTGGTCTCGACACAAATTTCACAGTCACATCGAAAACCGCAGATGATTTAGAA 72055
  864 PheGlyGluLeuTyrMetValLysTrpThrThrLeuLeuIleProProThrThrLeuLeu 883
72054 TTCGGTGAGCTTTACATTGTCAAATGGACAACTCTCTTGATCCCTCCGACGTCACTTCTT 71995
  884 IleIleAsnMetSerGly---CysAlaGlyPheSerAspAlaLeuAsnLysGlyTyrGlu 902
71994 ATAATCAACTTGGTCGGAGTTGTTGCTGGATTCTCTGACGCTCTTAACAAAGGTTATGAA 71935
```

```
                /codon_start=1
                /product="pherophorin like protein"
                /protein_id="CAB78858.1"
                /db_xref="GI:7268650"
                /translation="MVAGKVRVTMGFHKSPSTKKTKDMPSPLPLPPPPPPLKPPSSG
                SATTKPPINPSKPGFTRSFGVYPPRASAQVHATAAAASHNGVVSELRRQVEELRERERA
                LLKTENLEIADKNGEIDELRKETARLAEDNERLRREFDRSEEMRRECETREKEMEAEI
                VELRKLVSSESDDHALSVSQRFQGLMDVSAKSNLIRSLKRVGSLRNLPEPITNQENTN
                KSISSSGDADGDIYRKDEIESYSRSSNSEELTESSSLSTVRSRVPRVPKPPPKRSISL
                GDSTENRADPPPQKSIPPPPPPPPPPLLQQFPPPPSVSKAPPPPPPPPPPKSLSIASA
                KVRRVPEVVEFYHSLMERDSTNERRDSTGGGNAAAEAILANSNARDMIGEIENRSVYL
                LAIKTDVETQGDFIRELIKEVGNAAFSDIEDVVEFVKWLDDELSYLFKVCKFVVVSLK
                VDERAVLKHFEWPEQKADALREAAFCYFDLKKLIGEASRFREDPRQSSSSALKKMQAL
                FEKLEHGVYSLSRMRENAATKFKSFQIPVDWMLETGITSQIKLASVKLAMKYMKRVSA
                ELEAIEGGGPEEEELIVQGVRFAFRVHQFAGGFLAETMKAFEELRQKARSCHVQCQSQ
                THQHKLCFRSTPC"
exon            9577. .9906
                /gene="AT4g18560"
                /number=1
gene            9612. .11059
                /gene="AT4g18570"
gene            complement(join(9612. .9715,10483. .10747,10973. .11059))
                /gene="AT4g18570"
CDS             complement(join(9612. .9715,10483. .10747,10973. .11059))
                /gene="AT4g18570"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="CAB78859.1"
                /db_xref="GI:7268651"
                /translation="MSENAAFPTSLIKNLMKSPCVSTSVFICNQVLLPRRRFHRRWNL
                FSSLWSLFASTSDKTQPLPVLFSLSPTQCLNSSAAVVEEAEEELWKQTEVVVVVEEAV
                AEVEEEEAESISAVAVAEPEEGGLSGGGGGGGRGGEGMSLVFLVEGDL"
exon            complement(9612. .9715)
                /gene="AT4g18570"
                /number=1
intron          complement(9716. .10482)
                /gene="AT4g18570"
                /number=1
intron          9907. .9957
                /gene="AT4g18570"
                /number=1
exon            9958. .10809
                /gene="AT4g18570"
                /number=2
exon            complement(10483. .10747)
                /gene="AT4g18570"
                /number=2
intron          complement(10748. .10972)
                /gene="AT4g18570"
                /number=2
intron          10810. .10959
                /gene="AT4g18570"
                /number=2
exon            complement(10973. .11059)
                /gene="AT4g18570"
                /number=3
exon            10980. .11111
                /gene="AT4g18560"
                /number=3
intron          11112. .11159
                /gene="AT4g18560"
                /number=3
exon            11160. .11377
                /gene="AT4g18560"
                /number=4
intron          11378. .11462
                /gene="AT4g18560"
                /number=4
exon            11463. .11574
                /gene="AT4g18560"
                /number=5
intron          11575. .12309
                /gene="AT4g18560"
                /number=5
exon            12310. .12447
                /gene="AT4g18560"
                /number=6
intron          12448. .12540
                /gene="AT4g18560"
                /number=6
exon            12541. .12672
                /gene="AT4g18560"
                /number=7
gene            12991. .13662
                /gene="AT4g18580"
gene            complement(join(12991. .13077,13176. .13264,13382. .13507,
                13596. .13662))
                /gene="AT4g18580"
CDS             complement(join(12991. .13077,13176. .13264,13382. .13507,
                13596. .13662))
                /gene="AT4g18580"
                /note="contains EST gb:AI998801.1, AI138148"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="CAB78860.1"
                /db_xref="GI:7268652"
                /translation="MATLAPVHAVSVLHSRKRGGCNGNFVSFSGVPLPYLIRPRRKIP
                PVDTFDFYNDSDDESGEGDKGHSDDEYVVVDKAEAIMPNVKTEKPGAIQMMELSKDAV
                GDDESKAKDAGYSSDEWFVV"
exon            complement(12991. .13077)
                /gene="AT4g18580"
                /number=1
intron          complement(13078. .13175)
                /gene="AT4g18580"
                /number=1
exon            complement(13176. .13264)
                /gene="AT4g18580"
                /number=2
intron          complement(13265. .13381)
                /gene="AT4g18580"
                /number=2
exon            complement(13382. .13507)
                /gene="AT4g18580"
                /number=3
intron          complement(13508. .13595)
                /gene="AT4g18580"
                /number=3
exon            complement(13596. .13662)
                /gene="AT4g18580"
                /number=4
gene            14662. .20681
                /gene="AT4g18590"
CDS             join(14662. .14917,15014. .15052,16278. .16535,16652. .16752,
                17087. .17423,17520. .17854,18312. .18366,19447. .19527,
                19615. .20681)
                /gene="AT4g18590"
                /note="overlap to next genmodel
                similarity to 18.3K protein precursor, pollen, Zea mays,
                PIR2:JQ1107
                Contains Pollen proteins Ole e I family signature
                AA309-318
                contains EST gb:AI999674.1, AA721906"
                /codon_start=1
                /product="pollen-specific protein-like"
                /protein_id="CAB78861.1"
                /db_xref="GI:7268653"
                /translation="MDTSSPAAFVNGALLRRYIGQKVRAVIQVIRSDVGSVIGKSTDD
                QQIVVKGSPQPPLTTYLEVIGIAETDNTIRAEVWTNFGDSFDVQNYNELCKLANDRRK
                AMTDLQMEVEVDTNSSLQESLPKPQVMYRCKKCRRIVAIEENIVPHEPGKGEECFAWK
                KRSGNSEQVQCSSIFVEPMKWMQTIHDGMVEEKLLCFGCNGRLGYFNWAGMQCSCGAW
                LKITDGFNFMILYKTELCGSCELVVLFSNISNKENDYKREEVSKRRPESFNILKQMAS
                KAIFFFFVSAVCLGSLAGVAIADADDFDRFQIQGSVYCDTCRVQFVTRLSKFLEGAKV
                KLECRSRTNGTVTLTKEAVTDKTGSYRMEVTGDHEEEVCELVLVESPDSGCSDVSKEA
                YLRNAAKISLTANDGIVSHETRIVNPLGFMVQTPSAECPAAFKELGIVPDGLLFGWFG
                GQARMISETDGRSVTTHGAATCISGSFKEGYVESGTSAKARYAAAGSFGRIFSLGTGS
                SGKHLLCEPVSVVKLGTPEPVIFSSKTASVEDRSVVIVSLGKVLLCGLVSVVELEASE
```

```
ignment Scores:
ed. No.:                0              Length:         199075
ore:                    3926.00        Matches:        806
rcent Similarity:       68.69%         Conservative:   87
st Local Similarity:    62.00%         Mismatches:     81
ery Match:              74.92%         Indels:         331
                        8              Gaps:           16

-09-980-043-2 (1-978) x ATCHRIV49 (1-199075)

        1 MetMetGluSerGlyAlaProIleCysHisThrCysGlyGluGlnValGlyHisAspAla 20
94858 ATGATGGAGTCTAGGTCTCCCATCTGCAACACTTGTGGTGAAGAGATTGGTGTAAAATCA 94799

       21 AsnGlyGluLeuPheValAlaCysHisGluCysSerTyrProMetCysLysSerCysPhe 40
94798 AACGGAGAGTTCTTTGTGGCTTGTCATGAGTGTAGTTTCCCGATCTGCAAAGCTTGTCTT 94739

       41 GluPheGluIleAsnGluGlyArgLysValCysLeuArgCysGlySerProTyrAspIle 60
94738 GAGTATGAATTCAAAGAAGGTCGAAGAATTTGCTTGCGTTGCGGCAATCCTTACGATGGC 94680

       60 ------------------------------------------------------------ 60
94679 TACAATTATAAATTTACAAACTTTATTCGAAAGTTTGATTTTTTGGTTAAAGAGTATTTA 94620

       61 ---------------------------------AsnLeuLeuAspAspValGluLysLys 69
94619 TAACTTGTTTTGTTTTGTTTTGCTTTCATAGAGAATGTGTTTGATGATGTTGAGACAAAG 94560

       70 GlySerGlyAsnGlnSerThrMetAlaSerHisLeuAsnAspSer------ - -- --- 84
94559 ACATCTAAAACTCAATCCATTGTTCCAACACAGACCAATAACAC-TTCTCAGGTTTGTGT 94501

       84 ------------------------------------------------------------ 84
94500 GCTTAGCTAAGATTAGTTGCTTTTATGTCTTAATTAGATTTTACAAATTTGAGAACATTA 94441

       85 ------------GlnAspValGlyIleHisAlaArgHisIleSerSerValSerThrVal 100
94440 TCTTGACTTTTTCAGGATTCAGGGATTCATGCTAGACATATAAGTACAGTCTCAACAATA 94381

      101 AspSer------------------------------------------------------ 102
94380 GACAGTGGTAAGAACTAAGAAACAAAGCCAAAATCTTTGTATACATTTGGTGAGGTTATT 94321

      103 ---------------------------------------GluMetAsnAspGluTyrGl 109
94320 AATAGTTTAACTAACTTGGGGATTGTTGTTTCTTCCTGAAGAACTGAATGATGAATATGG 94261

      109 yAsnProIleTrpLysAsnArgValLysSerCys---------LysAspLysGluAsnLy 126
94260 CAATCCAATTTGGAAGAACAGAGTGGAGAGCTGGAAAGACAAGAAAGACAAGAAGAGCAA 94201

      126 sLysLysLysArgSerPro---------LysAlaGluThrGluProAlaGlnValProTh 143
94200 GAAGAAGAAGAAAGATCCAAAAGCAACAAAAGCTGAACAACATGAGGCTCAGATTCCTAC 94141

      143 rGluGlnGlnMetGluGluLysPro------------------------------ --- 151
94140 CCAACAGCACATGGAAGATACGCCACCGTAAGCACATCCTCGTACTAATCTAATAACATT 94081

      152 ------------------------------------------------------Se 152
94080 CACCAAAAAGTTTGAGACGTTGATTTGGTCTCTGCCGTTTGTATTCTCAGGAACACAGAA 94021

      152 rAlaGluAlaSerGluProLeuSerIleValTyrProIleProArgAsnLysLeuThrPr 172
94020 TCTGGT-GCTACAGATGTGCTTTCGGTTGTGATTCCTATCCCAAGGACAAAAATCACTTC 93962

      172 oTyrArgAlaValIleIleMetArgLeuValIleLeuGlyLeuPhePheHisPheArgIl 192
```

```
Db    93961 ATATAGGATTGTCA  ATCATGC  TGATCATCTTGGCTCTGTTTTTTAACTACCGTAT 93902

Qy      192 eThrAsnProValA pSerAlaPheGlyLeuTrpLeuThrSerValIleCysGluIleTr 212

Db    93901 CACGCATCCTGTCGATAGCGCTTA  TTAT  TAACATCTGTGATATGTGAGATTTG 93842

Qy      212 pPheAlaPheSerTrpValLeuAsp  hePro LysTrpAsnProValAsnArgGluTh 232

Db    93841 GTTTGCTGTTTCTT GGTGTTG AT  TCC TAAATGGTCTCCTATTAACCGAGAAAC 93782

Qy      232 rTyrIleGluArgLeuSerAlaArg Ty:      -----  -----  ------ ---- 241

Db    93781 TTACATCGACCGGT  TCG AAA TGTTTTATAACCGTTATATGCTTAATCT 93722

Qy      242 ------ -                          ----------  ---   ----G 242

Db    93721 GCGCAGTTACTA  TTCTTAA H   AT TA CATTTATCAAAACTATG AGATTCG 93662

Qy      242 luArgGlu yT  ---- -  - ---- -  AspPheTheValSerThrValAspP 262

Db    93661 AAACAGAAA GA  AATCATA  TA  TA  AGATTTCTTTGTTAGTA GTAGATC 93602

Qy      262 roLeuLysGluPr roLeuIleTh AlaAsnTh ValLeuSerIleLeuAlaValAspT 282

Db    93601 CATTAAAGGAGCCA  CTTTGATAACTGC AAA  ACGGTTCTTTCGATCCTCGCGCTTGATT 93542

Qy      282 yrProValAspLysValSerCysTyrValSer AspAspGlyAlaAlaMetLeuSerPheG 302

Db    93541 ATCCGGTGGATAAAGTCTCTTGCTATGTATCT GATGATGGTGCTGCAATGCTTTCGTTTG 93482

Qy      302 luSerLeuValGluThrAlaGluPheAlaArgLysTrpValProPheCysLysLysPheS 322

Db    93481 AGTCTTTGGTTGAGACAGCAGATTTTGCTAGGAAATGGGTACCTTTCTGCAAAAAGTACT 93422

Qy      322 erIleGluProArgAlaProGluPheTyrPheSerGlnLysIleAspTyrLeuLysAspL 342

Db    93421 CCATCGAGCCACGA  TCCCGAGTTTTA TTCTTCGCTTAAAATCGATTACTTGAGGGATA 93362

Qy      342 ysValGlnProSer heValLysGluArgArgAlaMetLys---------- ---- - 355

Db    93361 AAGTTCAACCTTCT TTGTGAAAGAACGTAGA CCATGAAAGTAAGTAAGCAAGGGCTTT 93302

Qy      355 --------  - ---  ---  ------ -----  ---------- 355

Db    93301 AAAACCTTAAATTAATCTCATATAGCGTTAAC GATGATTGAATCGAAAACTTAATGTCT 93242

Qy      356 -------  -----  -ArgAspTyrGluGluTyrLysValArgValAsnAlaLeuValA 370

Db    93241 AATTTCTCTTCTTG AGAGAGATTATGAAGAGTTTAAAATAAGAATGAATGCTTTAGTCG 93182

Qy      370 laLysAlaGlnLys hrProGluGluGlyTrpThrMetGlnAspGlyThrProTrpProG 390

Db    93181 CCAAGGCTGAAAAG CACCAGAAGAAGGATGGACAATGCAAGATGGAACATCTTGGCCCG 93122

Qy      390 lyAsnAsnThrArg spHisProGlyHisAspSerGlyLeu------  - ----  -- 403

Db    93121 GGAACAA ACTCTG  ACATCCCGGAT  GATTCAGGTTTGCTAGCTAAAACATTTGACC 93063

Qy      404 --      ----  --  --                  ---   -----  ---ProTrpG 406

Db    93062 TGAAATGCTCTGTC  TTTCTTGTTA  CAGTAATTTCTTTCTTGCAGGTTTTTCTTGG- 93004

Qy      406 luIle-LeuGlyAlaArgAspIleGluGlyAsnGluLeuProArgLeuValTyrValSer 425

Db    93003 --ATATAGCGGTGC CGCGACATTGAAGGAAATGAACTTCCAAGATTAGTTTACGTCTCT 92946

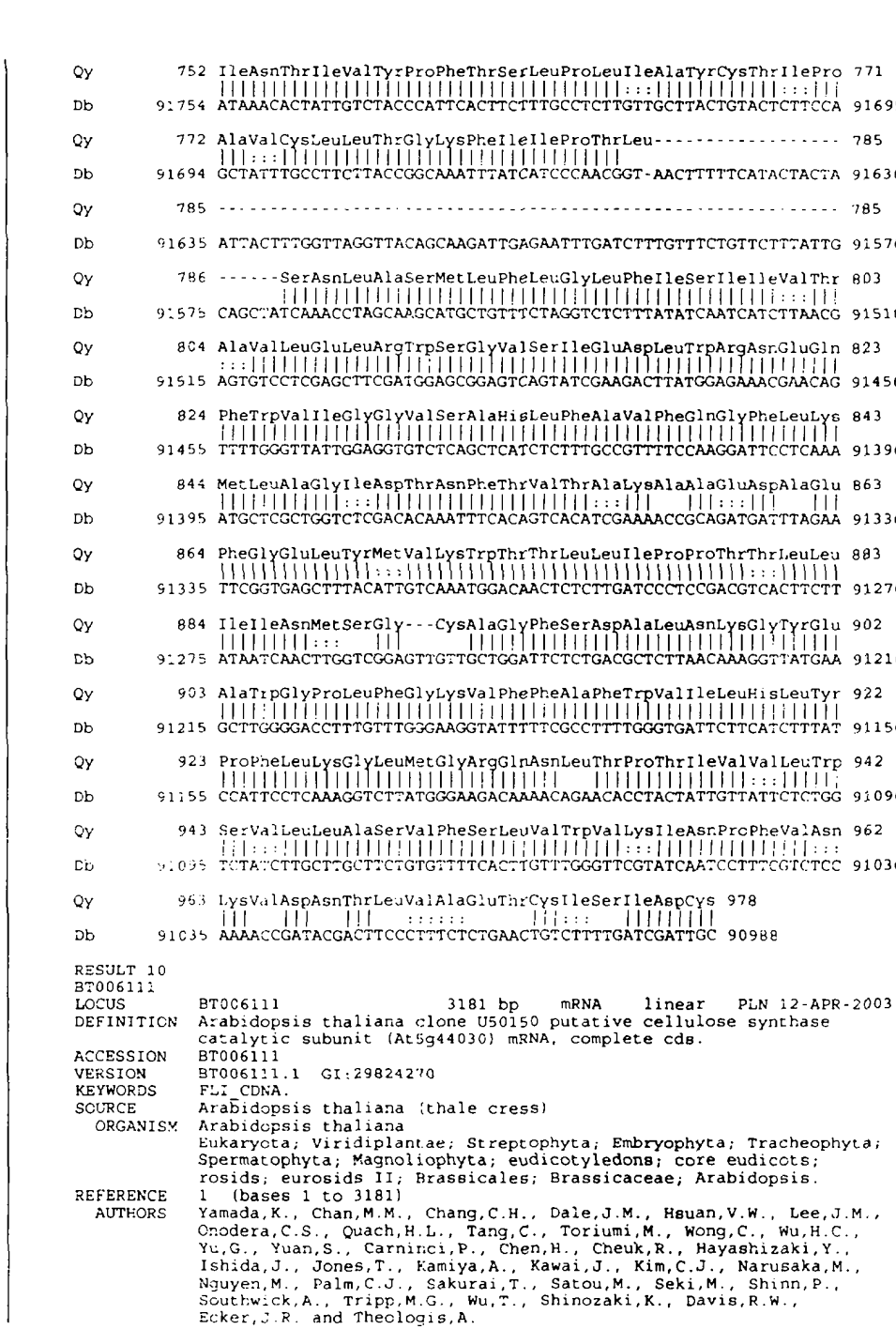
Qy      426 ArgGluLysArgPro GlyTyrGlnHisHisLysLysAlaGlyAlaGluAsnAlaLeu -- 444

Db    92945 AGAGAGAAGAGACC GGTTATCAGCATCACAAAAAGGCCGGGGCAGAGAACGCATTGGTA 92886

Qy      444 ------------  ----------  ----  ------------------  ------- 444

Db    92885 AACTTCAAAACTTTTTCTTTTATCGTTCAATTTAGCACAATCCTAATGATTTTGTTCTTT 92826
```





```
RESULT 10
BT006111
LOCUS       BT006111               3181 bp    mRNA    linear   PLN 12-APR-2003
DEFINITION  Arabidopsis thaliana clone U50150 putative cellulose synthase
            catalytic subunit (At5g44030) mRNA, complete cds.
ACCESSION   BT006111
VERSION     BT006111.1  GI:29824270
KEYWORDS    FLI_CDNA.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 3181)
  AUTHORS   Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M.,
            Onodera,C.S., Quach,H.L., Tang,C., Toriumi,M., Wong,C., Wu,H.C.,
            Yu,G., Yuan,S., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y.,
            Ishida,J., Jones,T., Kamiya,A., Kawai,J., Kim,C.J., Narusaka,M.,
            Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shinn,P.,
            Southwick,A., Tripp,M.G., Wu,T., Shinozaki,K., Davis,R.W.,
            Ecker,J.R. and Theologis,A.
```

```
TITLE     Arabidopsis Open Reading Frame (ORF) Clones
JOURNAL   Unpublished
FERENCE   2  (bases 1 to 3181)
AUTHORS   Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M.,
          Onodera,C.S., Quach,H.L., Tang,C.C., Toriumi,M., Wong,C., Wu,H.C.,
          Yu,G., Yuan,S., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y.,
          Ishida,J., Jones,T., Kamiya,A., Kawai,J., Kim,C.J., Narusaka,M.,
          Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shinn,P.,
          Southwick,A., Tripp,M.G., Wu,T., Shinozaki,K., Davis,R.W.,
          Ecker,J.R. and Theologis,A.
TITLE     Direct Submission
JOURNAL   Submitted (12-APR-2003) Plant Gene Expression Center, 800 Buchanan
          Street, Albany, CA 94710, USA
MMENT     The RIKEN Genomic Sciences Center (GSC) members carried out the
          collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN
          Arabidopsis  Full-Length  cDNA'): Seki,M., Narusaka,M.,  Ishida,J.,
           Satou,M.,  Kamiya,A., Sakurai,T., Carninci,P.,  Kawai, J.,
          Hayashizaki,Y.  and  Shinozaki,K.

          The Salk, Stanford, PGEC (SSP) Consortium members constructed and
          sequenced the pUNI (ORF) clones using the RAFL cDNAs: Yamada,K.,
          Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M.,
          Onodera,C.S.,  Quach,H.L., Tang,C.C., Toriumi,M., Wong,C., Wu,H.C.,
           Yu,G., Yuan,S., Chen,H., Cheuk,R., Jones,T., Kim,C.J., Nguyen,M.,
          Palm,C.J., Shinn,P., Southwick,A., Tripp,M.G., Wu,T., Davis,R.W.,
          Ecker,J.R. and Theologis,A.

          Yamada,K. (SSP/PGEC) and Seki, M. (RIKEN GSC) contributed equally
          to  this work.  Shinozaki,K. (RIKEN GSC) and Theologis,A. (SSP
          /PGEC)  contributed equally to this work as PIs.

          Annotation based on July 2002 version of the Arabidopsis genome
          submitted to Genbank.
ATURES              Location/Qualifiers
   source           1. .3181
                    /organism="Arabidopsis thaliana"
                    /mol_type="mRNA"
                    /db_xref="taxon:3702"
                    /chromosome="5"
                    /clone="U50150"
                    /note="This clone is in pUNI 51.
                    ecotype: Columbia"
   gene             1. .3181
                    /gene="At5g44030"
   CDS              1. .3150
                    /gene="At5g44030"
                    /codon_start=1
                    /evidence=experimental
                    /product="putative cellulose synthase catalytic subunit"
                    /protein_id="AAP04096.1"
                    /db_xref="GI:29824271"
                    /translation="MEPNTMASFDDEHRHSSFSAKICKVCGDEVKDDDNGQTFVACHV
                    CVYPVCKPCYEYERSNGNKCCPQCNTLYKRHKGSPKIAGDEENNGPDDSDDELNIKYR
                    QDGSSIHQNFAYGSENGDYNSKQQWRPNGRAFSSTGSVLGKDFEAERDGYTDAEWKER
                    VDKWKARQEKRGLVTKGEQTNEDKEDDEEEYLDAEARQPLWRKVPISSSKISPYRIVI
                    VLRLVILVFFFRFRILTPAKDAYPLWLISVICEIWFALSWILDQFPKWFPINRETYLD
                    RLSMRFERDGEKNKLAPVDVFVSTVDPLKEPPIITANTILSILAVDYPVNKVSCYVSD
                    DGASMLLFDTLSETSEFARRWVPFCKKYNVEPRAPEFYFSEKIDYLKDKVQTTFVKDR
                    RAMKREYEEFKVRINALVAKAQKKPEEGWVMQDGTPWPGNNTRDHPGMIQVYLGKEGA
                    FDIDGNELPRLVYVSREKRPGYAHHKKAGAMNAMVRVSAVLTNAPFMLNLDCDHYINN
                    SKAIRESMCFLMDPQLGKKLCYVQFPQRFDGIDLNDRYANRNIVFFDINMRGLDGIQG
                    PVYVGTGCVFNRPALYGYEPPVSEKRKKMTCDCWPSWICCCCGGGNRNHKSDSSKKKS
                    GIKSLFSKLKKKTKKKSDDKTMSSYSRKRSSTEAIFDLEDIEEGLEGYDELEKSSLMS
                    QKNFEKRFGMSPVFIASTLMENGGLPEATNTSSLIKEAIHVISCGYEEKTEWGKEIGW
                    IYGSVTEDILTGFRMHCRGWKSVYCMPKRPAFKGSAPINLSDRLHQVLRWALGSVEIF
                    FSRHCPLWYAWGGKLKILERLAYINTIVYPFTSIPLLAYCTIPAVCLLTGKFIIPTIN
                    NFASIWFLALFLSIIATAILELRWSGVSINDLWRNEQFWVIGGVSAHLFAVFQGLLKV
                    LFGVDTNFTVTSKGASDEADEFGDLYLFKWTTLLIPPTTLIILNMVGVVAGVSDAINN
                    GYGSWGPLFGKLFFAFWVIVHLYPFLKGLMGRQNRTPTIVVLWSILLASIFSLVWVRI
                    DPFLPKQTGPLLKQCGVDC"
   3'UTR            3151. .3181
                    /gene="At5g44030"
SE COUNT      900 a    654 c    748 g    879 t
```

```
ORIGIN

Alignment Scores:
Pred. No.:                1e-[illegible]    Length:              3181
Score:                    34[illegible]     Matches:             669
Percent Similarity:       75.[illegible]    Conservative:        119
Best Local Similarity:    64 [illegible]    Mismatches:          177
Query Match:              66.[illegible]    Indels:              76
DB:                       8                 Gaps:                15

US-09-980-043-2 (1-975) x US-[illegible]-[illegible]-1.rge

Qy          6 AlaProIleCysH...rpCysGly...GlnVa.GlyHisAspAlaAsnGlyGluLeuPhe 25
              ||  ||| ..|    |.|.|..  ..|'      |||  ..:  |||||||...  ...|
Db         58 GCTAAGATTTGCAA...CCTTCTGCTTA...AGTCTAAAGACGATGACAATGGTCAGACTTCT 117

Qy         26 ValAla.yHis.T...yyCys.....Thr.Gly...LysSerCysPheGluPheGluIleAsn 45
              ||  .'    |      |         |  .      .|:::|.:::  .  :::
Db        118 GTGAAGCTCATA...CTTCTA...CTTCTTCGCTATGAATATTAGCGTAATC 172

Qy         46 GluGlyArgLysVa...TyrLeuAla.....                  GlySerPro 57
              |||  |:      |  ::::|                          :|||||:.
Db        173 AACGGTAAGAAATG...TGCCCTCAATGAAACAC...TTTACAAACGCCACAAAGGCTCTCCA 237

Qy         58 ---- -    -  - ----          ---  TyrAspGluAsnLeuLeuAsp 64
                                              |||:::  ||:
Db        238 AAAATCGTCGGAGA...GAAGAAAATAATCTTC...ATGATTCAGATGATGAGTTGAATATC 297

Qy         65 AspValGluLysL, ..GlySerGlyAsn...SerThrMetAla---------- Ser 79
                  :::  |||||                    |:|
Db        298 AAATATCGCCAGGA...GGTTCCTTCAA...TACTAGAATTTTGCCTATGGATCGGAAAATGGA 357

Qy         80 HisLeuAsnAspSe...GlnAspVal.Gly...HisAlaArgHisIleSerSerValSerThr 99
                 |||    |.        :::  |||     ||||||  :::
Db        358 CACTACAACAGTAA...AGCAATGGC...AAAGGTCGAGCCTTTTCTTCCACCGGAAGT 417

Qy        100 Val --       Asp..SerGluMet   AsnAspGluTyrGlyAsnProIleTrpLysAsn 115
              ||:          :  |||          |||  |||  :::      ||:|.|
Db        418 GTGTTGGGGAAAGA...TTTGAAGCTGA...AGAGATGGCTACACAGATGCGGAGTGGAAAGAA 477

Qy        116 ArgValLysSerCy..LysAspLysGluAsnL..ysLysLysLysArgSerProLysAlaGlu 135
              |:|||          |.'  ::::::          :::|:|||                 :::
Db        478 CGTGTAGATAAATC...AAAGCGAGGCAA-          GAGAAGAGAGGTTTAGTGACTAAA 528

Qy        136 ThrGluProAlaGl..ValProThrGluGlnGl..MetGluGluLysProSerAlaGluAla 155
              |||             |||          |||||'                 ||:||||:
Db        529 GGAGAACAAACAAA...GAAGACAAAGAAGATGATG...AAGAAGAATACCTAGATGCTGAAGCA 588

Qy        156 SerGluProLeuSe..IleValTy..ProIleP..ArgAsnLysLeuThrProTyrArgAla 175
              :::| |||.        |||  ;      :::|:::::  ||||.;'|
Db        589 AGACAACCTCTCTG...AGAAAAGTCCC...AATCTC...TTCGAGCAAAATCAGTCCTTACCGCATC 648

Qy        176 ValIleIle-MetA..LeuValIleLeuGlyL..hePheHisPheArgIleThrAsnPro 195
              |||||:::::  | |'.|||  |||      '  ||||  ||||'|||,      :|
Db        649 GTGATCGCTCTTTG...CTTGTGATCCTTA...TTTCTTCCGTTTCCGTATCTTGACACCA 708

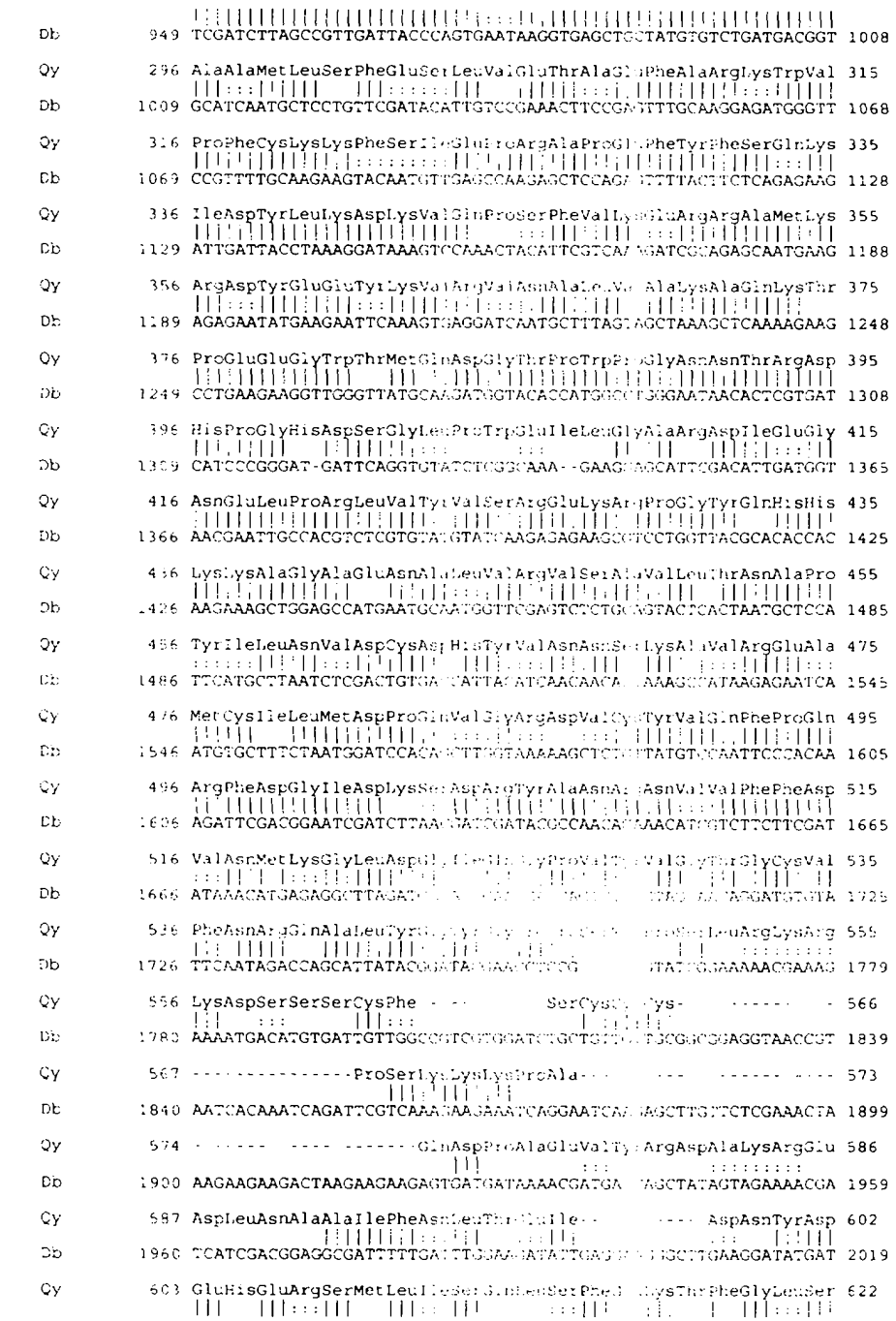
Qy        196 ValAspSerAlaP..GlyLeuTrpLeuThrSerValIleCysGluIleTrpPheAlaPhe 215
                  ||:::  ||||.|||,  ||.  ||||| ||||:|||||||
Db        709 GCCAAAGACGCATA...CCTCTCTGGCTTATCTCACTCATCTGCGAGATCTGGTTCGCTTTC 768

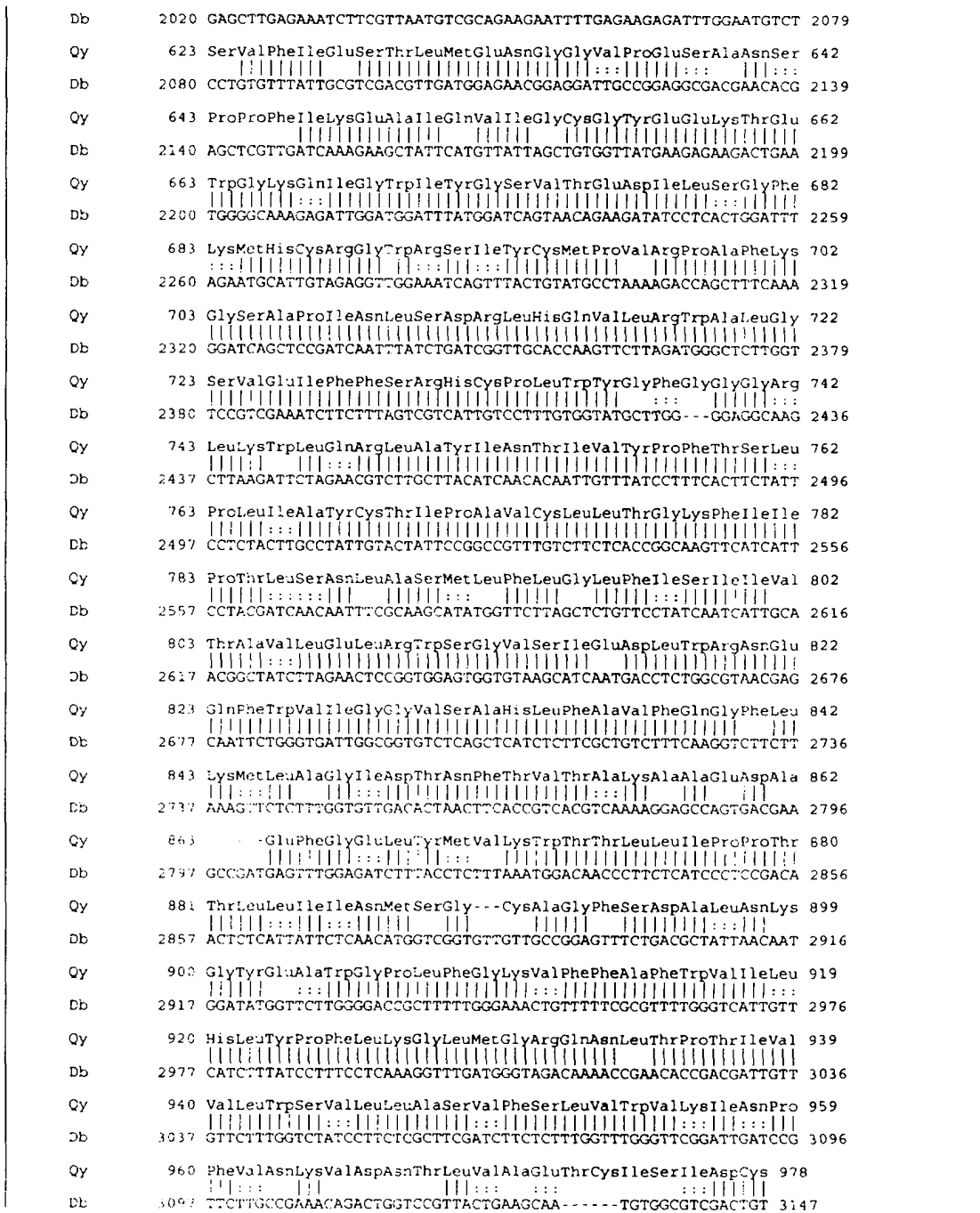
Qy        216 SerTrpValLeuAs..GlnPheProLysTrpAsnProValAsnArgGluThrTyrIleGlu 235
              :|||||:::|||| |||||.|||  |||   ||:::|||||||'|||||:::::
Db        769 TCTTGGATCCTTGA...CAGTTCCCGAAATGGTTTCCGATTAACCGAGAAACCTACCTTGAC 828

Qy        236 ArgLeuSe..AlaA..rTyrGluArgGluGlyGluProSerGlnLeuAlaGlyValAspPhe 255
              ||||||||   ||:::||||||:::||||,|   :::::||||||   |||||,
Db        829 CGTCTCTCTATGAGATTTGAACGGGATGGGGAAAAGAACAAGCTTGCACCGGTTGATGTG 888

Qy        256 PheValSerThrVa..AspProLeuLysGluProProLeuIleThrAlaAsnThrValLeu 275
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        889 TTTGTCAGTACAGT...GATCCGCTCAAGGAACCTCCCCATAATCACAGCCAACACAATTCTT 948

Qy        276 SerIleLeuAlaValAspTyrProValAspLysValSerCysTyrValSerAspAspGly 295
```





```
SULT 11
005710
CUS       BT005710                3377 bp    mRNA    linear   PLN 15 MAR-2003
FINITION  Arabidopsis thaliana clone RAFL15-30-K05 (R50150) putative
          cellulose synthase catalytic subunit (At5g44030) mRNA, complete
          cds.
CESSION   BT005710
RSION     BT005710.1  GI:28973610
YWORDS    FLI_CDNA.
URCE      Arabidopsis thaliana (thale cress)
ORGANISM  Arabidopsis thaliana
          Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
          Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
          rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
FERENCE   1  (bases 1 to 3377)
AUTHORS   Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M.,
          Onodera,C.S., Quach,H.L., Tang,C., Toriumi,M., Wong,C., Wu,H.C.,
          Yu,G., Yuan,S., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y.,
          Ishida,J., Jones,T., Kamiya,A., Kawai,J., Kim,C.J., Narusaka,M.,
          Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shinn,P.,
          Southwick,A., Tripp,M.G., Wu,T., Shinozaki,K., Davis,R.W.,
          Ecker,J.R. and Theologis,A.
TITLE     Arabidopsis Full Length cDNA Clones
JOURNAL   Unpublished
FERENCE   2  (bases 1 to 3377)
AUTHORS   Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M.,
          Onodera,C.S., Quach,H.L., Tang,C.C., Toriumi,M., Wong,C., Wu,H.C.,
          Yu,G., Yuan,S., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y.,
          Ishida,J., Jones,T., Kamiya,A., Kawai,J., Kim,C.J., Narusaka,M.,
          Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shinn,P.,
          Southwick,A., Tripp,M.G., Wu,T., Shinozaki,K., Davis,R.W.,
          Ecker,J.R. and Theologis,A.
TITLE     Direct Submission
JOURNAL   Submitted (15-MAR-2003) Plant Gene Expression Center, 800 Buchanan
          Street, Albany, CA 94710, USA
MMENT     RIKEN Genomic Sciences Center (GSC) members carried out the
          collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN
          Arabidopsis Full-Length cDNA'): Seki,M., Narusaka,M., Ishida,J.,
          Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J.,
          Hayashizaki,Y. and Shinozaki,K.

          The Salk, Stanford, PGEC (SSP) Consortium members carried out the
          sequencing and annotation of the RAFL cDNAs: Yamada,K., Chan,M.M.,
          Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M., Onodera,C.S.,
          Quach,H.L., Tang,C.C., Toriumi,M., Wong,C., Wu,H.C., Yu,G.,
          Yuan,S., Chen,H., Cheuk,R., Jones,T., Kim,C.J., Nguyen,M.,
          Palm,C.J., Shinn,P., Southwick,A., Tripp,M.G., Wu,T., Davis,R.W.,
          Ecker,J.R. and Theologis,A.

          Yamada,K. (SSP/PGEC) and Seki,M. (RIKEN GSC) contributed equally to
          this work. Shinozaki,K. (RIKEN GSC) and Theologis,A. (SSP/PGEC)
          contributed equally to this work as PIs.

          Annotation based on July 2002 version of the Arabidopsis genome
          submitted to Genbank.
ATURES              Location/Qualifiers
   source           1. .3377
                    /organism="Arabidopsis thaliana"
                    /mol_type="mRNA"
                    /db_xref="taxon:3702"
                    /chromosome="5"
                    /clone="RAFL15-30-K05 (R50150)"
                    /note="This clone is in a modified pBluescript vector2
                    (lambda PS) as a BamHI/XhoI insert.
                    ecotype: Columbia"
   gene             1. .3377
                    /gene="At5g44030"
   5'UTR            1. .71
                    /gene="At5g44030"
   misc_difference  2. .3
                    /gene="At5g44030"
                    /note="compared to genomic sequence"
                    /replace="aaa"
   CDS              72. .3[illegible]
                    /gene="At5g44030"
                    /codon_start=1
                    /evidence=experimental
                    /product="putative cellulose synthase catalytic subunit"
                    /protein_id="AAO64[illegible].1"
                    /db_xref="GI:28973611"
                    /translation="MSPNTMASFDDFHRHSSFSAKICKVCGDEVKDDNGQTFVACHV
                    CVYPVCKPCYEYERSNGNKCCPQCNTLYKRHKGSPKIAGDEENNGPDDSDDELNIKYP
                    CDGSSIHLNPAYGSENGDYKSKQQWRPNGRAFSSTGSVLGKDFEAERDGYTLAEWKER
                    VDKWKARQNKRGLVCKGQTNEDKEDDEEEYLDAEARQPLWRKVPISSSKISPYRIVI
                    VLRLVILCFFFRFRILTPAKDAYPLWLISVICEIWFALSWILDQFPKWFPINRETYLD
                    RLSMRFE[illegible]GFKNKLAPVDVFVSTVDPLKEPPIITANTILSILAVDYPVNKVSCYVSD
                    DGASMLTFTLSETSEFARRWVPFCKKYNVEPRAPEFYFSEKIDYLKDKVQTTFVKDR
                    RAMKREY[illegible]PFVRINAC[illegible]AKAHKPPEGWVMQDGTPWPGNNTRDHPGMIQVYLGKEGA
                    RDIDGN[illegible]GAMNAMVRVSAVLTNAPFMLNLDCDHYINN
                    [illegible]RFDGIDLNDRYANRNIVFFDINMRGLDGIQG
                    PVYV[illegible]Pr PAMTCDCWPSWICCWGGSNRNHKSDSSKKKS
                    [illegible]KRSSTEAIFDLSDIEEGLEGYDEIEKSSLMS
                    QKNFEKR[illegible]MSFVFIAT[illegible]MEN[illegible]PSATNTSSLIKEAIHVISCGYEEKTEWGKEIGW
                    IYGSVTE[illegible]DGFRMHCPGWKSVYCMPKRPAFKGSAPINLSDRLHQVLRWALGSVEIF
                    FSKHCPLAYAWGGKLKIFERLAYINTIVYPFTSIPLLAYCTIPAVCLLTGKFIIPTIN
                    NFASIWFIALFLSIIATAILELRWSGVSINDLWRNEQFWVIGGVSAHLFAVFQGLLKV
                    LFGVDTNFTVTSKGASDEADEFADLYLFKWTTLLIPPTTLIILNMVGVVAGVSDAINN
                    GYGSWGPLFGKLFFAFWVIVHLYPFLKGLMGRQNRTPTIVVLWSILLASIFSLVWVRI
                    DPFLPFQTPLLKQVGVF"
   3'UTR            3202. .3377
                    /gene="At5g44030"
BASE COUNT      960 a    694 c    780 g    943 t
ORIGIN

Alignment Scores:
Pred. No.:                1.1e-293        Length:          3377
Score:                    3465.00         Matches:         669
Percent Similarity:       75.86%          Conservative:    119
Best Local Similarity:    64.34%          Mismatches:      177
Query Match:              66.04%          Indels:          76
DB:                       8               Gaps:            15

US-09-980-043-2 (1-978) x BT005710 (1-3377)

Qy          6 AlaProIleCysH[illegible]ThrCysGlyGluGlnValGlyHisAspAlaAsnGlyGluLeuPhe 45
              ||| ||. ||    |||||[illegible]| |   |||  ||||[illegible]   |||
Db        129 GCTAAGATTTGCAAC[illegible]TCTGTGGCGATGAGGTCAAAGACGATGACAATGGTCAGACTTTT 188

Qy         26 ValAlaCysHisGl[illegible]ysSerTyrProMe[illegible]CysLysSerCysPheGluPheGluIleAsn 45
              ||| |||| ||   ||  |||. |[illegible] ||||  |||[illegible]|||[illegible]|||   :::
Db        189 GTGGCGTGTTCACGT[illegible]TGCGTTTACC[illegible]GGTTTG[illegible]AAACCTTGCTATGAATATGAGCGTAGC 248

Qy         46 GluGlyArgLysVa[illegible]ysLeuArgCys---      --------  ------GlySerPro 87
                 ||    ||  |||  ||||  |                                  ||||||||
Db        249 AACGGTAA[illegible]AAATC[illegible]GCCCTCAA[illegible]GAACA[illegible]TTTTACAAACGCCACAAAGGGTCTCCA 308

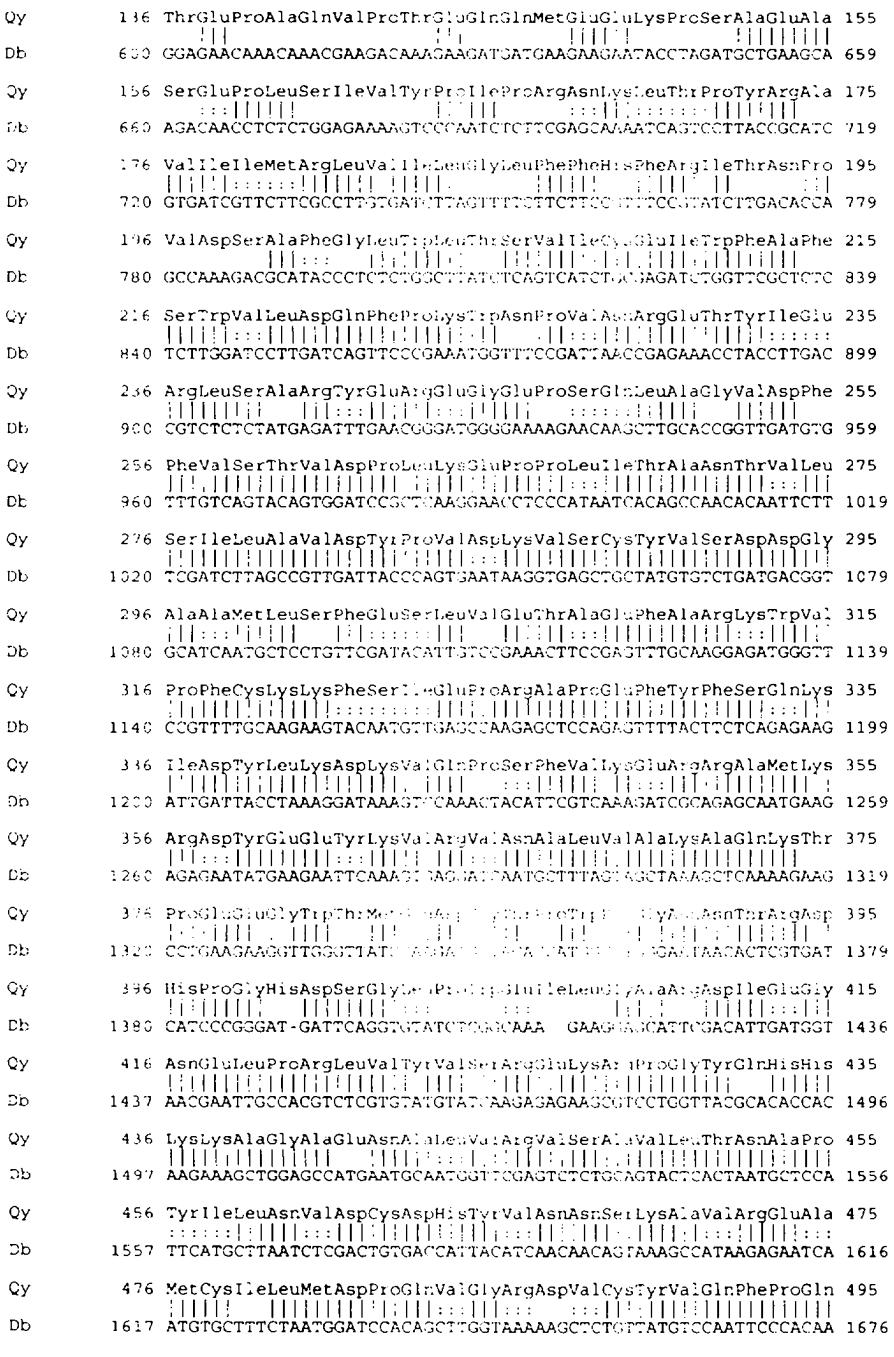
Qy         88 --  -             -------  -      ---  TyrAspGluAsnLeuLeuAsp 114
                                                     ||||:::  ||.
Db        309 AAAATG[illegible]GAGA[illegible]AAGAAAACAATG[illegible]TGATGATTCAGATGATGAGTTGAATATC 368

Qy         65 AspVal[illegible]LysLy[illegible]GlySerGlyAsnGluSer[illegible]ThrMetAla----      --      Ser 79
                 :::  .|  ||                                ||
Db        369 AAATATC[illegible]AGGA[illegible]GTTCCTCCATTTACCA[illegible]AATTTTGCCTATGGATCGGAAAATGGA 428

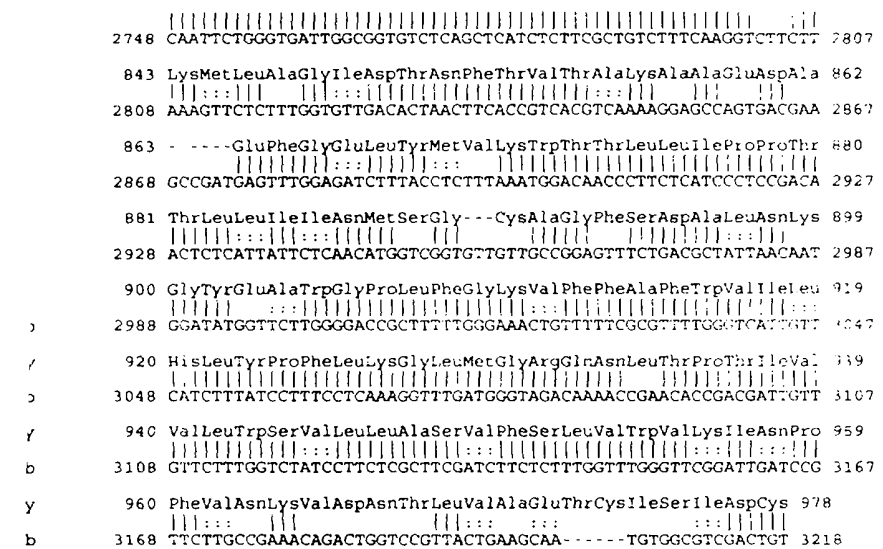
Qy         80 HisLeuAsnAspSer[illegible]lnAspValGlyIleHisAlaArgHisIleSerSerValSerThr 99
                 ||          |||  ||  :::  |||    !.||||      :::
Db        429 GACTACAACAGTAA[illegible]CAGCAATGGCGTCCAAATGGTCGAGCCTTTTCTTCCACCGGAAGT 488

Qy        100 Val----    --AspSerGluMet---AsnAspGluTyrGlyAsnProIleTrpLysAsn 115
                 :||       |:  |||        |||  |||  :::      ||.|||
Db        489 GTGTTGGGGAAAGA[illegible]TTTGAAGCTGAGAGAGATGGCTACACAGATGCGGAGTGGAAAGAA 548

Qy        116 ArgValLysSerCysLysAspLysGluAsnLysLysLysLysArgSerProLysAlaGlu 135
                 ||||||          |||   ::::::          :::||||||              :::
Db        549 CGTGTAGATAAATGGAAAGCGAGGCAA-----  -GAGAAGAGAGGTTTAGTGACTAAA 599
```







```
ESULT 12
F458083
OCUS       AF458083               3168 bp    mRNA    linear   PLN 04-FEB-2003
EFINITION  Arabidopsis thaliana cellulose synthase (IRX5) mRNA, complete cds.
CCESSION   AF458083
ERSION     AF458083.1  GI:27462650
EYWORDS    .
OURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
EFERENCE   1  (bases 1 to 3168)
  AUTHORS   Taylor,N.G., Howells,R.M., Huttly,A.K., Vickers,K. and Turner,S.R.
  TITLE     Interactions among three distinct CesA proteins essential for
            cellulose synthesis
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 100 (3), 1450-1455 (2003)
   PUBMED   12538856
EFERENCE   2  (bases 1 to 3168)
  AUTHORS   Taylor,N.G., Vickers,K. and Turner,S.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-DEC-2001) Biological Sciences, University of
            Manchester, Oxford Road, Manchester M13 9PT, UK
EATURES             Location/Qualifiers
     source         1. .3168
                    /organism="Arabidopsis thaliana"
                    /mol_type="mRNA"
                    /db_xref="taxon:3702"
     gene           1. .3168
                    /gene="IRX5"
     CDS            1. .3168
                    /gene="IRX5"
                    /codon_start=1
                    /product="cellulose synthase"
                    /protein_id="AAO15532.1"
                    /db_xref="GI:27462651"
                    /translation="MEPNTMASFDDEHRHSSFSAKICKVCGDEVKDDDNGQTFVACHV
                    CVYPVCKPCYEYERSNGNKCCPQCNTLYKRHKGSPKIAGDEENNGPDDSDDELNIKYR
                    QDGSSIHQNFAYGSENGDYNSKQQCRPNGRAFSSTGSVLGKDFEAERDGYTDAEWKER
                    VDKWKARQEKRGLVTKGEQTNEDKEDDEEEELLDAEARQPLWRKVPISSSKISPYRIV
                    IVLRLVILVFFFRFRILTPAKDAYPLWLISVICEIWFALSWILDQFPKWFPINRETYL
                    DRLSMRFERDGEKNKLAPVDVFVSTVDPLKEPPIITANTILSILAVDYPVNKVSCYVS
```



```
                    DPGASMGPDTLSETSEFAKRWVPFCKKYNVEPRAPEFYFSEKIDYLKDKVQTTFVKD
                    RRAMKREYEEFKVRINALVAKAQKKIEEGWVMQDGTPWPGNNTRDHPGMIQVYLGKEG
                    AFDIDGNELPRLVYVSREKFAFQYHHKKAGAMNAMVRVSAVLTNAPFMLNLDCDHYIN
                    NSKAIREAMCFLMDPQLGKEICYVQFPQRFDGIDRHDRYANRNIVFFDINMRGLDGIQ
                    GPVYVGTGCVFNRPALYGFEPPVSEKRKKMTCDCWPSWICCCCGGGNRNHHKSKSDS
                    SSKKKSGIKSLLSKIKKKKKKSDQKTMSSYSRKRSATEAIFDLRDIEEGLEGYDELE
                    KSSLMSQKNFEKRFGMSPVFIASTLMENGGLPEATNTSSLIKEAIHVISCGYEEKTEW
                    GKEIGWIYGSVTEDILTGFKMHCRGWKSVYCMPKRPAFKGSAPINLSDRLHQVLRWAL
                    GSVEIFFSRHCPLWYAWGGKLKWLERLAYINTIVYPFTSIPLLAYCTIPAVCLLTGKF
                    IIPTINNFASIWFLALFLSIIATGILELRWSGVSINDLWRNEQFWVIGGVSAHLFAVF
                    QGLLKVLAGVDTNFTVTSKGADEAEEFGDLYLFKWTTLLIPPTTLIILNMVGVVAGV
                    SDAINNGYQSWGPLFGKLFFAFWVIVHLYPFLKGLMGRQNRTPTIVVLWSILLASIFS
                    LWWVRIDPFLPKQTGPDLEQCGVNC"
BASE COUNT      907 a    645 c    741 g    874 t
ORIGIN
```



```
Alignment Scores:
Pred. No.:                                  3168
Score:                                      666
Percent Similarity:                         122
Best Local Similarity:                      177
Query Match:                                81
DB:                                         15

US-09-980-043-2 (1 978) x AF458083 (1 3168)
```



```
Qy          6 AlaProIleCysHisThrCysGlyGluGlnValGlyHisAspAlaAsnGlyGluLeuPhe 25
Db         98 GCTAAGATTTGCAAAGTCTGTGGCGATGAGGTCAAAGACGATGACAATGGTCAGACTTTT 157

Qy         26 ValAlaCysHisGluCysSerTyrProValCysLysSerCysPheGluPheGluIleAsn 45
Db        118 GTGGCGTGTCACGTATGCGTTTACCCATTTGCAAACCTTGCTATGAATATGAGCGTAGC 177

Qy         46 GluGlyArgLysValCysLeuArgCys----------------------GlySerPro 57
Db        178 AACGGTAACAAATGTTGCCCTCAATTAAACACTCTTTACAAACGCCACAAAGGCTCTCAA 237

Qy         58 ---------------------------------------TyrAspGluAsnLeuLeuAsp 64
Db        238 AAAATCGCCGGAGATGAAGAAAACAATTGTCTTGATGATTCAGATGATGAGTTGAATATC 297

Qy         65 AspValGluLysLysGlySerGlyAsnHisSerThrMetAla---------------Ser 79
Db        298 AAATATCGCAGGATGGTTCCTCCATTCACGAAATTTGCCTATGGATCGGAAAATGGA 357

Qy         80 HisLeuAsnAspSerGlnAspValGlyLeuHisAlaArgHisIleSerSerValSerThr 99
Db        358 GACTACAATAGTAACAGCAATGGCCTCAAATGGCCGGCCTTTTCTTCCACCGGAAGT 417

Qy        100 Val---------AspSerGluMet-AsnAspGluTyrGlyAsnProIleTrpLysAsn 115
Db        418 GTGTTGGAAAAGAGTTTGAACCTGAGAAGATGGCTACACAGATGCAGAGTGGAAAGAA 477

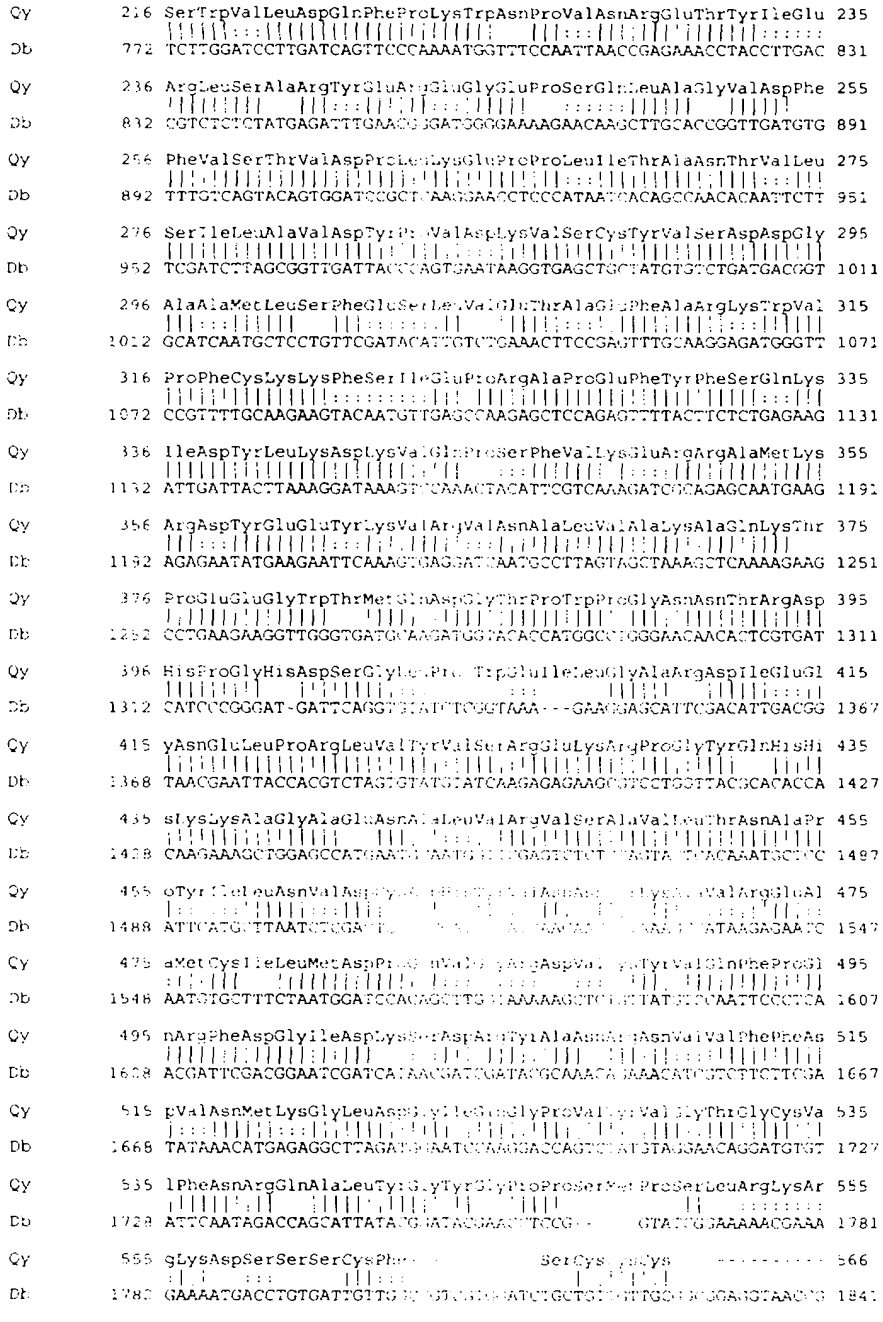
Qy        116 ArgValLysSerCysLysAspLysGlnArgLysLysLysArgSerProLysAlaGlu 135
Db        478 CGTGTAGATAAATGCAAAGCGAGCAAGAAGAAGAGGTTTAGTGACTAAAGGAGAACAA 537

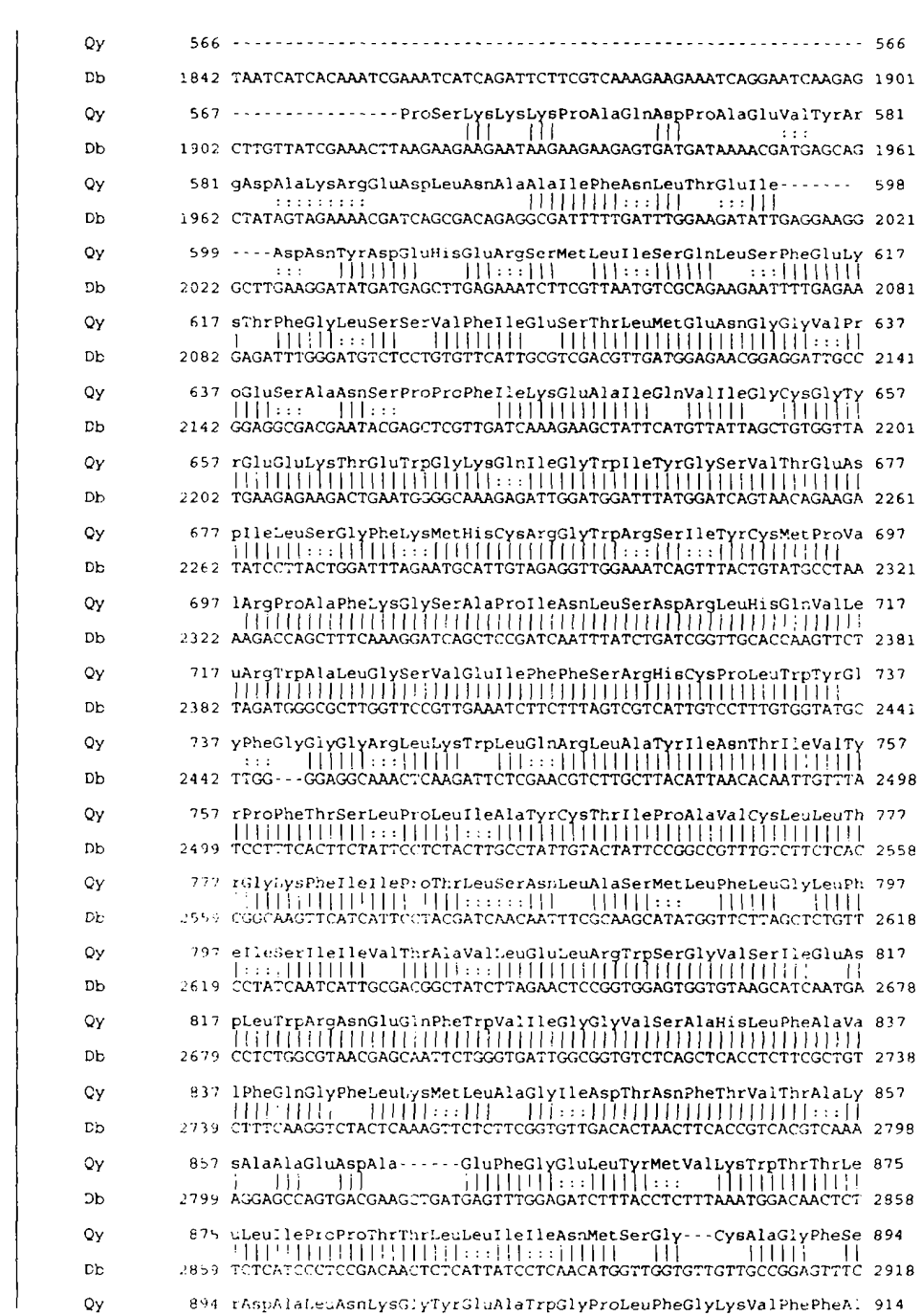
Qy        136 ThrGluProAlaGlnValProThrGluGlnGlnMetGluGluLysProSerAlaGluAla 155
Db        538 ACAAACGAAGACAAAGAAGATGATGAAGAA-----GAAGAATTACTAGATGCTGAAGCA 591

Qy        156 SerGluProLeuSerIleValTyrProIleProArgAsnLysLeuThrProTyrArgAla 175
Db        592 AGACAACCTCTCTGCAGAAAAGTCCCAATCTCTTCGAGCAAAATCAGTCCTTACCGTATC 651

Qy        176 ValIleIleMetArgLeuValIleLeuGlyLeuPhePheHisPheArgIleThrAsnPro 195
Db        652 GTGATCGTTCTTCGTCTTGTGATCTTAGTTTTCTTCTTCCGTTTCCGTATCTTGACACCA 711

Qy        196 ValAspSerAlaPheGlyLeuTrpLeuThrSerValIleCysGluIleTrpPheAlaPhe 215
Db        712 GCCAAAGACGCATACCCTCTCTGGCTTATCTCAGTCATCTGCGAGATCTGGTTCGCACTC 771
```





```
      |||||||:::|||   ||||||   :::|||||||||':||||||||.:::||.||||'
   2919 TGACGCTATTAACAATGGATATGGTTCTTGGGGACCGCTTTTTGGGAAACTGTTTTTCGC 2978

    914 aPheTrpValIleLeuHisLeuTyrProPheLeuLysGlyLeuMetGlyArgGlnAsnLe 934
        ||||||||||||||:::|||||||||||||||||||||||||||||||||||||||||||'
   2979 GTTTTGGGTCATTGTTCATCTTTATCCTTTCCTCAAAGGTTTGATGGGTAGACAAAACCG 3038

    934 uThrProThrIleValValLeuTrpSerValLeuLeuAlaSerValPheSerLeuValTr 954
        |||||||||||||||||||||||||:::|||||||||||:::||||||||||:||||
   3039 AACACCGACGATTGTTGTTCTTTGGTCTATCCTTCTCGCTTCGATCTTCTCTTTGGTTTG 3098

    954 pValLysIleAsnProPheValAsnLysValAspAsnThrLeuValAlaGluThrCysIl 974
        ||||:::|||:::||||||:::   |||           ||:::   :::
   3099 GGTTCGGATTGATCCGTTCTTGCCGAAACAGACTGGTCCGTTACTGAAGCAA------TG 3152

    974 eSerIleAspCys 978
        :::|||||||
   3153 TGGCGTCGACTGT 3165
```

```
SULT 13
003237
)CUS        AP003237             162776 bp    DNA     linear   PLN 29 MAR 2003
FINITION    Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1,
            PAC clone:P0046E05.
CESSION     AP003237 BA000010
RSION       AP003237.2  GI:15146360
'YWORDS     .
)URCE       Oryza sativa (japonica cultivar-group)
 ORGANISM   Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
:FERENCE    1
 AUTHORS    Sasaki,T., Matsumoto,T., Yamamoto,K., Sakata,K., Baba,T.,
            Katayose,Y., Wu,J., Niimura,Y., Cheng,Z., Nagamura,Y.,
            Antonio,B.A., Kanamori,H., Hosokawa,S., Masukawa,M., Arikawa,K.,
            Chiden,Y., Hayashi,M., Okamoto,M., Ando,T., Aoki,H., Arita,K.,
            Hamada,M., Harada,C., Hijishita,S., Honda,M., Ichikawa,Y.,
            Idonuma,A., Iijima,M., Ikeda,M., Ikeno,M., Itoh,S., Itoh,T.,
            Itoh,Y., Itoh,Y., Iwabuchi,A., Kamiya,K., Karasawa,W., Katagiri,S.,
            Kikuta,A., Kobayashi,N., Kono,I., Machita,K., Maehara,T.,
            Mizuno,H., Mizubayashi,T., Mukai,Y., Nagasaki,H., Nakashima,M.,
            Nakama,Y., Nakamichi,Y., Nakamura,M., Namiki,N., Negishi,M.,
            Ohta,I., Ono,N., Saji,S., Sakai,K., Shibata,M., Shimokawa,T.,
            Shomura,A., Song,J., Takazaki,Y., Terasawa,K., Tsuji,K., Waki,K.,
            Yamagata,H., Yamane,H., Yoshiki,S., Yoshihara,R., Yukawa,K.,
            Zhong,H., Iwama,H., Endo,T., Ito,H., Hahn,J.H., Kim,H.I., Eun,M.Y.,
            Yano,M., Jiang,J. and Gojobori,T.
 TITLE      The genome sequence and structure of rice chromosome 1
 JOURNAL    Nature 420 (6913), 312-316 (2002)
 MEDLINE    22337376
  PUBMED    12447438
EFERENCE    2  (bases 1 to 162776)
 AUTHORS    Sasaki,T., Matsumoto,T. and Yamamoto,K.
 TITLE      Direct Submission
 JOURNAL    Submitted (19-FEB-2001) Takuji Sasaki, National Institute of
            Agrobiological Sciences, Rice Genome Research Program; Kannondai
            2-1-2, Tsukuba, Ibaraki 305-8602, Japan
            (E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
            Tel:81-298-38-7441, Fax:81-298-38-7468)
OMMENT      On Aug 9, 2001 this sequence version replaced gi:13027267.
            Genes were predicted from the integrated results of the following:
            GENSCAN1.0, BLASTN2.0, BLASTX2.0 as well as SplicePredictor
            (October 1998 version). The genomic sequence was searched against
            NCBI NonRedundant Protein database, nr
            (ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA sequence database at
            RGP. Protein homologies of the coding regions were searched against
            NCBI NonRedundant Protein database with BLASTP2.0. ESTs represent
            the identified cDNA sequences using BLASTN 2.0 with the
            corresponding DDBJ accession no. and RGP clone ID.
            A gene with identity or significant homology to a protein is
            classified based on the protein name to indicate the homology level
            such as same name, 'putative-' and '-like protein'. A gene without
            significant homology to any protein but with EST homology (covering
            almost the entire length of partial sequence) is classified as an
            'unknown' protein. A gene predicted with a gene prediction program
            is classified as a 'hypothetical' protein.
            The orientation of the sequence is from T7 to SP6 of the PAC clone.
            This sequence of P0046E05 clone has an overlap with P0481E12 (DDBJ:
            AP003076) clone at the position 1 to 1344 of 3' end. The sequence
            of this clone starts at the position 160,220 of P0481E12. Detailed
            information on overlap and assembly quality together with
            annotation of this entry is available at
            http://rgp.dna.affrc.go.jp/publication.html.
FEATURES    Location/Qualifiers
     source     1..162776
                /organism="Oryza sativa (japonica cultivar-group)"
                /mol_type="genomic DNA"
                /cultivar="Nipponbare"
                /db_xref="taxon:39947"
                /chromosome="1"
                /clone="P0046E05"
     gene       complement(join(491..[illegible],782..838,1548..1647,1804..1908,
                1963..2109,2358..2424,2469..3233,3750..3866,3928..4034))
                /gene="P0046E05.1"
     CDS        complement(join(491..[illegible],782..838,1548..1647,1804..1908,
                1963..2109,2358..2424,2469..3233,3750..3866,3928..4034))
                /gene="P0046E05.1"
                /note="hypothetical protein"
                /codon_start=1
                /protein_id="BAB67896.1"
                /db_xref="GI:15623837"
                /translation="MSGRGQALLRAPRLAVLYSWFVVAAAGGVAGEGRLAAPGERDSS
                RRELALAGALHGRPLHSAALGALLSRQRGRPPLPSPAALPTAALTTEDEQRCGRPNGG
                PWAGWQRASAAAGGDRGGEPPTQWRVCVGVAVTVELGSARQPPHSRLPPTTTVEEKSR
                FKRMRDPLIRTIQSYIMGDQMQFEQJTGRQPGNVKEIKNRSRQYRACMSLSVLASRQA
                SIPAFLISLRPQKQAQHSHNYSSQPEKNGARKISAALRYKPAQARRPRERRGDVKSG
                PFVSTCSRSFGRYTAHTRTRTRTKATEVSLLISEEEAEEDERVINPSRKQFFDKPKQ
                LPLSRMIFIDGWIERPVDPYRPARLPSDSSSFFASNPIAKSGARNGLRIPGIGILMV
                ACLH"
     gene       join(5029..5176,5301..5396)
                /gene="P0046E05.2"
     CDS        join(5029..5176,5301..5396)
                /gene="P0046E05.2"
                /note="hypothetical protein"
                /codon_start=1
                /protein_id="BAB67897.1"
                /db_xref="GI:15623838"
                /translation="MVVKKKQPHHTFTTTTAARAKTRTSDLRNDAERAAAAWPGAKTW
                PSPEEKRFLSHSAATTIQKPSTRGHRSGASPLVSHHEGARTPTRSSSIHAIGLDIERV
                REKNGSPLPHTSLLDVAIFYRQRRHRTLLEKRFFMAFGKYFSWRL"
     gene       join(5998..6122,6184..6253)
                /gene="P0046E05.3"
     CDS        join(5998..6122,6184..6253)
                /gene="P0046E05.3"
                /note="hypothetical protein"
                /codon_start=1
                /protein_id="BAB67898.1"
                /db_xref="GI:15623839"
                /translation="MTGSPRRARAPAGRLDPPPSYHDDDSDHHHHHDDDDDHSDRSS
                RREEAAVTAVAMASGRRFFR"
     gene       complement(join(8862..8972,9711..9837,10015..10116))
                /gene="P0046E05.4"
     CDS        complement(join(8862..8972,9711..9837,10015..10116))
                /gene="P0046E05.4"
                /note="hypothetical protein"
                /codon_start=1
                /protein_id="BAE67899.1"
                /db_xref="GI:15623840"
                /translation="MERERGGEAGEEEACARQVFGVFFHGVLRGFEAKPACCDGCGPD
                ARWEVRCSDLAQEFFVFHMFTATGTAQHGKSFDHNWTSSSTFATT"
     gene       join(10402..10564,10665..10774,10872..10944,11052..11199,
                11931..12197,12298..12640,12729..12866,13453..13578,
                13690..13902,13982..14228,14311..14510,14749..14955,
                15048..15632)
                /gene="P0046E05.5"
```

```
CDS             join(10402. .10564,10665. .10774,10872. .10944,11052. .11199,
                11931. .12197,12298. .12640,12729. .12866,13453. .13578,
                13690. .13902,13982. .14228,14311. .14510,14749. .14955,
                15048. .15632)
                /gene="P0046E05.5"
                /note="contains ESTs D24711(R2417),AU031890(R2417)"
                /codon_start=1
                /product="putative cellulose synthase"
                /protein_id="BAB67900.1"
                /db_xref="GI:15623841"
                /translation="MESGVFPCAACGDDAHAACRACSYALCKACLDEDAAEGRTTCAR
                CGGEYGAPDPAHGQGAVVEEEVEESHEPAAGGVRERVTMASQLSDHQDEGVHARTMST
                HARTISSVSGVGSELNDRSCKPIWKNRVESWKEKKKEKKASAKKAAAKAQAPPVEEQI
                MDEKDLTDAYEPLSRIIPISKNKLTFYRAVIIMRLVVLGLFFHYRITNPVYSAFGLWM
                TSVICEIWFGFSWILDQFPKWCPINRETYVDRLIARYGDGEDSGLAPVDFFVSTVDPL
                KEPPLITANTVLSILAVDYPVEKISCYVSDDGSAMLTFESLAETAEFARRWVPFCKKY
                SIEPRAPEFYFSQKIDYLKDKIHPSFVKERRAMKRDYEEYKVRINALVAKAQKTPEEG
                WIMQDGTPWPGNNPRDHPGMIQVFLGETGARDFDGNELPRLVYVSREKRPGYQHHKKA
                GAMNALVRVSAVLTNAPYILNLDCDHYVNNSKAVREAMCFMMDPSVGRDVCYVQFPQR
                FDGIDRSDRYANRNVVFFDVNMKGLDGLQGPVYVGTGCCFYRQALYGYGPPSLPALPK
                SSVCSWCCCCCPKKKAEKSEKEMHRDSRREDLESAIPNLREIDNYDEYERSMLISQMS
                FEKSFGLSSVFIESTLMENGGVPESANPSTLIKEAIHVISCGYEEKTEWGKEVLRWAL
                GSVEIFLSRHCPLWYGYGGGRLKWLQRLSYINTIVYPFTSLPLIAYCCLPAICLLTGK
                FIIPTLSNAATIWFLGLFISIIVTSVLELRWSGIGIEDWWRNEQFWVIGGVSAHLFAV
                FQGILKMIAGLDTNFTVTAKATDDTEFGELYVFKWTTVLIPPTSILVLNLVGVVAGFS
                DALNSGYESWGPLFGKVFFAMWVIMHLYPFLKGLMGRQNRTPTIVVLWSVLLASVFSL
                LWVKIDPFIGSSETTTNSCANFDC"
gene            complement(join(17113. .17220,17766. .17888,18013. .18118,
                19257. .20727,21600. .21867))
                /gene="P0046E05.6"
CDS             complement(join(17113. .17220,17766. .17888,18013. .18118,
                19257. .20727,21600. .21867))
                /gene="P0046E05.6"
                /note="hypothetical protein
                similar to Arabidopsis thaliana F28A21.170"
                /codon_start=1
                /protein_id="BAB67901.1"
                /db_xref="GI:15623842"
                /translation="MDKQRQVFSKIGIHGSVQRFNWIKSCQSDLEVLGTIKVQKFHVA
                TTAKQDLDSRSSTTPFMAQGLLLSALLQELGDNATPLQRLPLHKGILPKATAPPASHP
                SVRMPTNGPGSRHGPVHPALSQLKTSETARAPAHPPFETFLGVSPCLLFASASKSMAA
                ATTTTVLATHDLLLLGLLLVAISFAGAAPLDPEQLVALRAIGLRPHRLDPCDDAAGAV
                GVVAASCDAGVPFRRVTSLVMANCSATTSVSAGALEALAFSLRALSFSDCPAAPPRLL
                PPEQLAAGLRAFSCVASLHRLSAVWLSHLENLTELTVADTPLATGSPSELAVVVSHMG
                HLTSLTLSNANLSGFLPHHWHCPNLTHLDLSGNRITGAIPDTLTLLSAITHLNLSSND
                LNGNIPTSIGDLISLTTIDLSNNSISGRIPDTVSTLPELEVLNLGSNRLNGSIPQFLS
                EMRGLKELNLEGNDFDGMVPFTAKFVSRLRVFRAAGNPKLCYNRSVLSAEIAIGVAPC
                DKYGFPVMAPPATATARSPSADYDLGAKDGDADDGADARGGPSAAVLGVAIGLSCLA
                FAVILLVCLCKVCRVYMPSSLLIPLCKEIPDAIPSFLAPPCGTNTQGPSTSRCQPR
                AIRMSRTAIMYPDVPLSLSSPFPRNVPATAIVPPLQGSRIRYIGYSSRDIIEYCDV
                LSVLMYLGK"
LTR             complement(28[illegible])
                /note="3' LTR"
gene            join(27694. .28956,29047. .30944)
                /gene="P0046E05.7"
CDS             join(27694. .28956,29047. .30944)
                /gene="P0046E05.7"
                /note="hypothetical protein
                similar to Oryza sativa chromosome 6, P0535G04"
                /codon_start=1
                /protein_id="BAB67902.1"
                /db_xref="GI:15623843"
                /translation="MARGSALLDGSVLPPSRIVSEPQAGLPRRFMPESATGREIVTLG
                EGRPAPDYPERSVFFLPFAMASLVPPFSSFFMDVLEFYDLQMAHLTPNAVMTLAIFAH
                LCEMFIGVRPSLRLFRWFFTVQSVSPPSVVGGCYFCQPRGPVLNRYIPCALRKKWDDWK
                SDWFYTPLADEARLRLPSCPPAQASSWRAPVDLGLGYDAVLDRLAGLRSQGLTGAMVY
                GDYLRRRIAPLQRRARGAWEYTGSEDFTRTHQGVRWDWAPEDFKILVQRVLNLNSVEA
                ALIPQGILPLCSVPDRASILTIMMAVGASEERAPKGHDGAGGSRRGEQFTPRGGRASG
                SRDRGPGSSRPADARGKRKQGGTPPPSPPRGGGAARVSSRRPEGAAPTSQPEGERKKK
                RPRKMGETEPSRGNLISPPKWSFNRPPRSEIPSRPSRHSKSGQSEAEDPAAVEALRRE
                SDRREAADRLREAEEAAQEAARARQAEETAREEAARARQAGEAAREEVGFRQDEAMAT
                SEAARDEAAGASLGPTPSGDAQATTSGAAGDEAAGASLGPTPSGDAQDQPGPGDIPES
                GTSIGGPNRVASSPRRLCPTPSIAPLNAEPLLQALAAANTTVLDGLSAQVEVLQAERA
```

```
                ELDAAWARVEEGRRSVEAMVEVGRKAHRRHVSELETRKKVLAEIAKEVEEERGAALIA
                TTVMNEAQDTLRLQYGSWEAELRKKLDAAQGVLDAAAAREQRAAETEAASRQREEALE

Alignment Scores:
Pred. No.:              1.36e-285       Length:         162776
Score:                  3401.50         Matches:        762
Percent Similarity:     48.07%          Conservative:   85
Best Local Similarity:  43.25%          Mismatches:     119
Query Match:            64.91%          Indels:         799
DB:                     8               Gaps:           24

US-09-980-043-2 (1-978) x AP003257 (1-162776)

Qy          1 MetMetGluSerGlyAlaProIleCysHisThrCysGlyGluGlnValGlyHisAspAla 20
              ||||||||||||||||   |||   |||      ||||||:::            ||||||
Db      10399 ATGATGGAGTCGGGGGTCCCGCCCTGCGCCGCGTGCGGGGAC------ -----GACGCG 10446

Qy         21 AsnGlyGluLeuPheValAlaCysHisGluCysSerTyrProMetCysLysSerCysPhe 40
              :::                 |||||      |||||||||   :::||||||:::|||
Db      10447 CAC-----------GCCGCCTGCCGCGCCTGCAGCTACGCGCTCTGCAAGGCCTGCCTC 10494

Qy         41 GluPheGluIleAsnGluGlyArgLysValCysLeuArgCysGlySerProTyr------ 58
              :::   :::      |||||||||      |||   |||||||||      |||
Db      10495 GACGAGGACGCCGCCGAGGGACGCACCACCTGCGCGCGCTGCGGCGGGGAGTACGGCGCG 10554

Qy         58 -------- ---------------------------------------------------- 58

Db      10555 CCCGACCCAGGTGCCCCCGCGCGCTGCGGGTTACTGTTTTGCCCTATGCGGGATTCGATC 10614

Qy         59 ---------AspGlu-----------AsnLeuLeuAsp-AspValGluLysLysGlySe 71
                        ||||||           :::   :::|||   |||          ||
Db      10615 GATGCGGGGGATGAAGTAGCAGTTTCTGACGAGGTTGATCTCTGTCACAGCGCATGGCCA 10674

Qy         71 rGly---------------------------------------------------------- 72
               |||
Db      10675 GGGGGCGGTCGTGGAGGAGGAAGTGGAAGAGAGCCACGAGCCGGCGGCCGGCGGTGTTCG 10734

Qy         73 -AsnGlnSerThrMetAlaSerHisLeuAsnAsp------------------------- 83
                 :::   ||||||||||||   |||:::|||
Db      10735 CGAGAGGGTCACCATGGCCAGCCAACTCAGCGATCACCAGGTGAATAGCCCTCTCCATAA 10794

Qy         83 ---------------------------------------------------------- 83

Db      10795 ACTCGTGCAATTCTGGGTAGAACAGCTTGAGCGCTCCGACGTTGTTGACAATGTTGTGTG 10854

Qy         84 ----------------------------SerGlnAsp-ValGlyIleHisAlaArgH 93
                                          |||||  :::     |||||||||
Db      10855 GGATTTTGTGGGAACAGGATGAAGGAGTTCATGCCAGGACTATGAGCACCCATGCCAGGA 10914

Qy         94 isIleSerSerValSerThrValAsp---------------------------------- 101
               |:|||:|||||||||   |||
Db      10915 CAATTAGCAGTGTCTCTGGAGTTGG-GAGTGGTAAGCAACTCGTCTCGTTGTCTGTCATA 10973

Qy        101 ---------------------------------------------------------- 101

Db      10974 TTTTTCCTTGCATTATTCCAACTGAGTACTGATATATATATATATATATATATTTTCA 11033

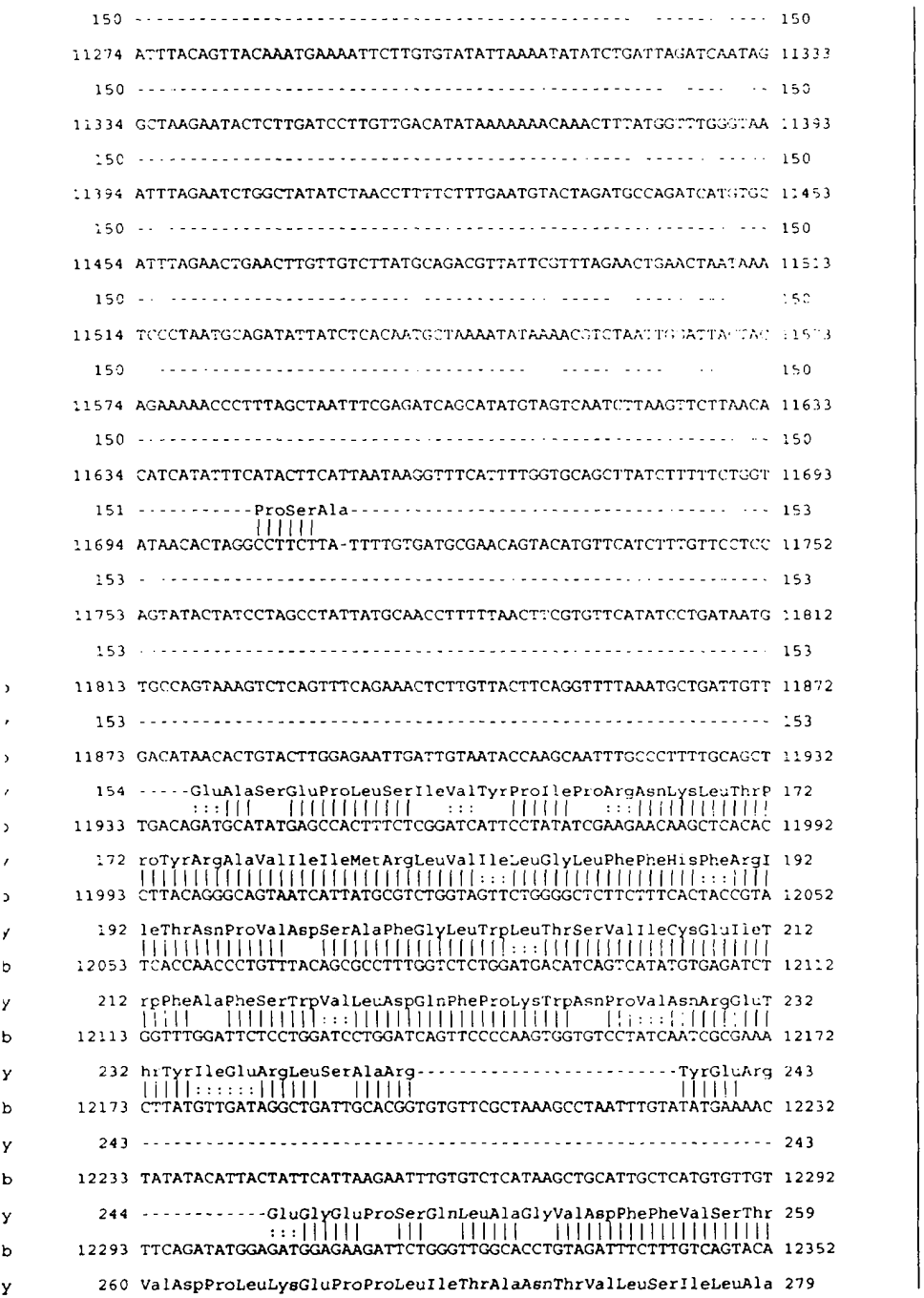
Qy        102 -------------SerGluMetAsnAspGluTyrGlyAsnProIleTrpLysAsnArgV 117
                          |||:::|||||||||   |||   |||||||||||||||||||
Db      11034 CTTGTTATGGATTTGAAGAATTGAATGATGAATCTGGGAAGCCCATCTGGAAGAACAGGG 11093

Qy        117 alLysSerCysLysAspLysGluAsnLysLysLysLysArgSerProLysAlaGluThrG 137
              ||:::|||   |||:::|||:::   :::||||||      :::   ||||||    :
Db      11094 TGGAGAGCTGGAAGGAGAAAAAGAAAGAGAAGAAAGCCTCAGCAAAGAAAGCTGCAGCAA 11153

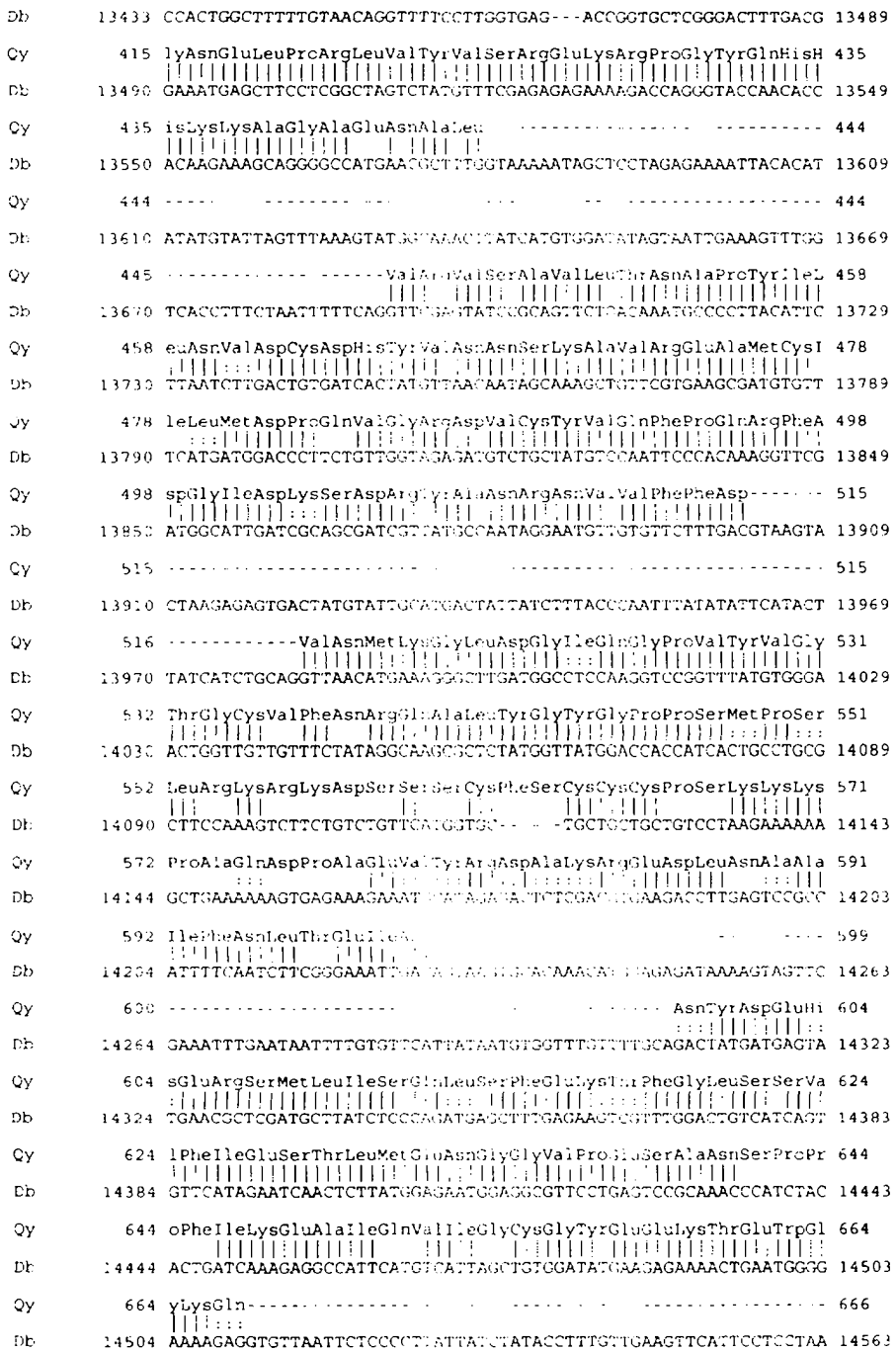
Qy        137 luProAlaGlnValProThrGluGlnGln---MetGluGluLys---------------- 150
              ::          |||   |||:::|||   |||:::||||||
Db      11154 AAGCTCAAGCTCCACCTGTTGAAGAACAGATTATGGATGAAAAAGAGTAAGGAACATGGA 11213

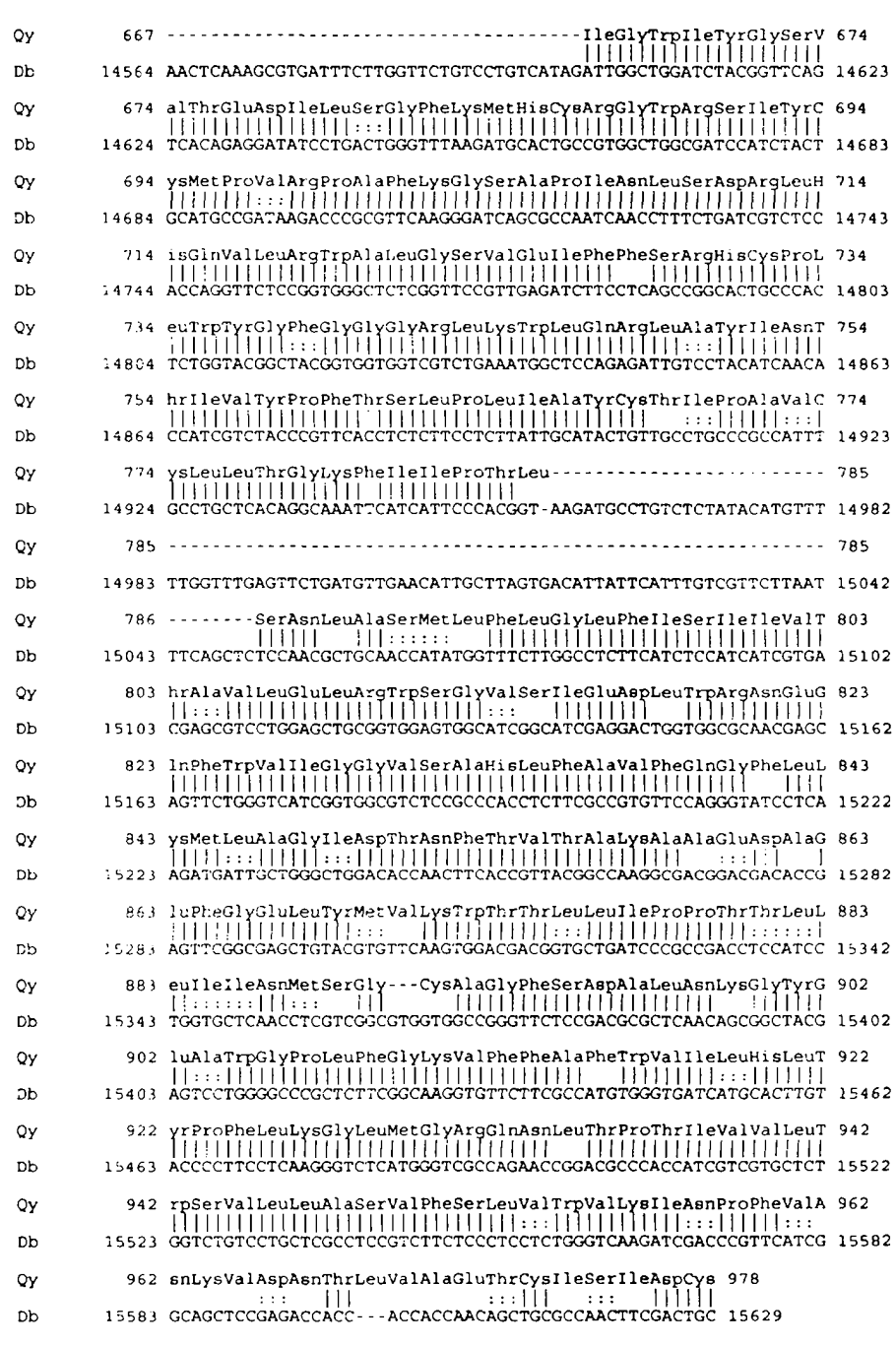
Qy        150 ---------------------------------------------------------- 150

Db      11214 ACCATTTCATAGATCCATGTGATCCATCAAAAACTTAACATTTCTTGATTCATATACATT 11273
```









```
SULT 14
081534
CUS       AF081534                3395 bp    mRNA    linear   PLN 29-NOV-2000
FINITION  Populus x canescens cellulose synthase (cell) mRNA, complete cds.
CESSION   AF081534
RSION     AF081534.1  GI:3511284
YWORDS    .
URCE      Populus x canescens
ORGANISM  Populus x canescens
          Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
          Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
          rosids; eurosids I; Malpighiales; Salicaceae; Populus.
FERENCE   1  (bases 1 to 3395)
AUTHORS   Wang,H. and Loopstra,C.A.
TITLE     Cloning and Characterization of a Cellulose Synthase cDNA
          (Accession No. AF081534) from Xylem of Hybrid Poplar (PGR98-179)
JOURNAL   Plant Physiol. 118 (3), 1101 (1998)
FERENCE   2  (bases 1 to 3395)
AUTHORS   Wang,H., Loopstra,C.A. and No,E.
TITLE     Direct Submission
JOURNAL   Submitted (30-JUL-1998) Forest Science, Texas A&M University,
          College Station, TX 77843, USA
ATURES              Location/Qualifiers
    source          1. .3395
                    /organism="Populus x canescens"
                    /mol_type="mRNA"
                    /strain="717"
                    /db_xref="taxon:80863"
                    /tissue_type="xylem"
    gene            1. .3395
                    /gene="cell"
    CDS             27. .3155
                    /gene="cell"
                    /codon_start=1
                    /product="cellulose synthase"
                    /protein_id="AAC78476.1"
                    /db_xref="GI:3511285"
                    /translation="MAGLVTGSSQTLHAKDELMPPTRQSATSKKCRVCGDEIGVKEDG
                    EVFVACHVCGFPVCRPCYEYERSEGNQSCPQCNTRYKRHKGCPRVPGDNDDEDANPDD
                    FDDEFQIKHHDHDESNQKNVFSHTEIEHYNEQEMHPIRPAFSSAGSVAGKDLEGDKEG
                    YSNAEWQERVEKWKVRQEKRGLVSKDEGGNDQGEEDEYLMAEARQPLWRKIPIPSSRI
                    NPYRIVIVLRLIILCFFFRFRILTPASDAYALWLISVICEVWFGLSWILDQFPKWNPI
                    ERETYLDRLSMRFEREGEPNRLGPVDVFVSTVDPLKEPPIITANTVLSILSVDYPVDK
                    VSCYVSDDGASMLLFDSLAETAEFARKWVPFCKKHNIEPRAPEFYFTQKIDYLKDKVH
                    PNFVKERRAMKREYEEFKVRINALVSKAQKKPEEGWVMQDGTPWPGNITRDHPGMIQV
                    YLGSEGALDVEGKELPRLVYVSREKRPGYNHHKKAGAMNALILVSAVLTNAPFMLNLD
                    CDHYINNSKAVREAMCFLMDPQLGKKLCYVQFPQRFDGIDRHDRYANRNVVFFDINMK
                    GLDGVQGPVYVGTGCVFNRQSLYGYDPPVSEKRPKMTCDCWPSWCCCCFGGSRKKSKK
                    KGQRSLLGGLYPMKKKMMGKKYTRKASAPVFDLEEIEEGLEGYEELEKSSLMSQKSLE
                    KRFGQSPVFIASTLMENGGVPEGTNSQSHIKEAIHVISCGYEEKTEWGKEVGWIYGSV
                    TEDILTGFKMHCRGWRSVYCSPKRPAFKGSAPINLSDRLHQVLRWALGSIEIFLSHHC
                    PLWYGYGGKLKLLERLAYINTIVYPFTSIPLLAYCTTPAVCLLTGKFIIPTLNNLASI
                    WFPGPFHLNHSNICVGVGWSGVSIQDLGRNEQFWVIGGVSGHLFAVFQGFFKVLGGVD
                    TNFTVTSKSADDAEFGELYLFKWTTLLIPPTTLIILNMVGVVAGVSARINNGYGSWGP
                    LFGKLFFAFWVIVHLYPFLKGLMGKQNRTPTIVVLWSVLLASIFSLIWVRIDPFLPKQ
                    TGPILKQCGVEC"
ASE COUNT      916 a    677 c    830 g    972 t
RIGIN

lignment Scores:
red. No.:              1.39e-287       Length:         3395
core:                  3396.00         Matches:        662
ercent Similarity:     76.76%          Conservative:   124
est Local Similarity:  64.65%          Mismatches:     174
uery Match:            64.81%          Indels:         68
B:                     8               Gaps:           17

S-09-980-043-2 (1-978) x AF081534 (1-3395)

y          9 CysHisThrCysGlyGluGlnValGlyHisAspAlaAsnGlyGluLeuPheValAlaCys 28
             |||      |||||:::::::::|||          :::||||||:::|||||||||||||
b        117 TGTAGAGTTTGTGGGGATGAGATTGGAGTTAAGGAAGATGGAGAGGTGTTTGTTGCTTGT 176

Qy        29 HisGluCysSerTyrProMetCys--SerCysPheGluPheGluIleAsnGluGlyArg 48
             ||   |||   :::'|:::  ||:::          :::||:::|||   :::'||||
Db       177 CATGTGTGTGGCTTTCCTGTTTGTA-TGCTTGTTATGAGTATGAGAGGAGTGAAGGCAAC 236

Qy        49 LysValCysLeuArgCysGlySerProTyr--  ----------      ----- 68
             :::  |||   :  ||'  :::  ||:
Db       237 CAGTCCTGTCCTCA-TGCAACACTT-ATA-AA-GTCACAAAGGTTGTCCTAGAGTTCCT 296

Qy        69 --   ---   AspGluAsnLeu-   LeuAspAspValGluLysLys-----      -- 89
                 ||  ||:::      :  '|||   :::   :::
Db       297 GGAGATAATGACGA-GAGGATGCCAACTTTGA-GATTTTGACGACGAATTTCAGATTAAG 356

Qy        90 ---------  --GlySerGlyAsnGlnSerThrMetAlaSer-------------- -His 82
                          ||  |||       :::   ||                    |||
Db       357 CATCATGATCATGATGAATCCAATCAGAAAAATGTTTTTAGTCACACGGAAATTGAGCAC 416

Qy        81 LeuAsnAspSer-GluAspValGly-IleHisAlaArgHisIleSer----  SerValSer 98
             :||:::  ::      :                  |||      '||||:::::
Db       417 TACAATGAA-AGA-AATGCA-----A-----GCCTTTTCGTCAGCAGGAAGTGTTGCT 476

Qy        99 ThrVal-Asp-Ser-G-MetAsnAsp---  ---CysAsnProIleTrpLysAsnArgVal 117
                 :|  |:  :::   ::    :  ||:    |:::::   |||:
Db       477 GGTAAGGATCTTGAA-GCGACAAAGA-GAA-AAA-CAATGCAGAATGGCAA-AGAGGTG 536

Qy       118 LysSerCysLysAspLysGluAsnLysLysLysLysArgSerProLysAlaGluThrGlu 137
             :::      ||   :::::      .:||:||      :::::   |||
Db       537 GAGAAATGCAAAGTTAGGCAA---GAAAAGAGAGGTTTGGTGAGCAAACGATGAA 587

Qy       138 ProAlaGlnValPro-ThrGluGlnGlnMet--GluGluLysProSerAlaGluAlaSerGlu 157
                          ||:::     ::|||         |||||||||   :::
Db       588 GGAGGAAATGATCAA-GGAGAAGAA---GATGAGTACCTTATGGCTGAAGCCAGGCAA 641

Qy       158 ProLeuSerIleValTyrProIleProArgAsnLysLeuThrProTyrArgAlaValIle 177
             ||'||         .||||:'||     :::::::    |||||||''   ||'|||
Db       642 CCACTATGCAGAAAAATCCCAATTC-TTCCAGCAGAATCAACCCGTATCGAATTGTCATT 701

Qy       178 IleMetArgLeuValIleLeuGlyLeuPheHisPheArgIleThrArgProValAsp 197
             ::::::|||:|:: |||||     '|||.    |.||||||     ||
Db       702 GTCCTTCGACTTATCATTCTTTGCTTCTTTTTCGTTTTCGGATCTTAACTCCAGCATCT 761

Qy       198 SerAlaPheGlyLeuTrpLeuThrSerValIleCysGluIleTrpPheAlaPheSerTrp 217
                 |||:::  .|||||'|||  |||.:'||.: ||||:::||||.!      |||||;
Db       762 GATGCTTATGCATTGTGGCTTATCTCTGTAATATGTGAGGTATGGTTTGGCCTCTCCTGG 821

Qy       218 ValLeuAspGlnPheProLysTrpAsnProValAsnArgGluThrTyrIleGluArgLeu 237
             :::|||||:'||||:'|||||| .'|||:..  :|'||||||||||:::::|||'|
Db       822 ATCCTTGACCAGTTCCCAAAATGGAACCCAATGAACGTGAAACTTATCTCGATCGCCTA 881

Qy       238 SerAlaArgTyrGluArgGluGlyGluProSerGlnLeuAlaGlyValAspPhePheVal 257
             |||  |||:::'||||:|||||||||' |:::::|||      ||||||  |||||
Db       882 TCCATGAGGTTTGAGCGTGAGGGTGAGCCCAATCAGTCTGGGCCCAGTTGATGTGTTTGTG 941

Qy       258 SerThrValAspProLeuLysGluProProLeuIleThrAlaAsnThrValLeuSerIle 277
             ||||'|||||'|||  |||:.|||.:.||| :::||:'||||.|||||.|||||||||||'
Db       942 AGTACTGTGGATC-C-TCAAGGAAC-AC-AATCATAACTGCAAATACAGTCCTTTCAATC 1001

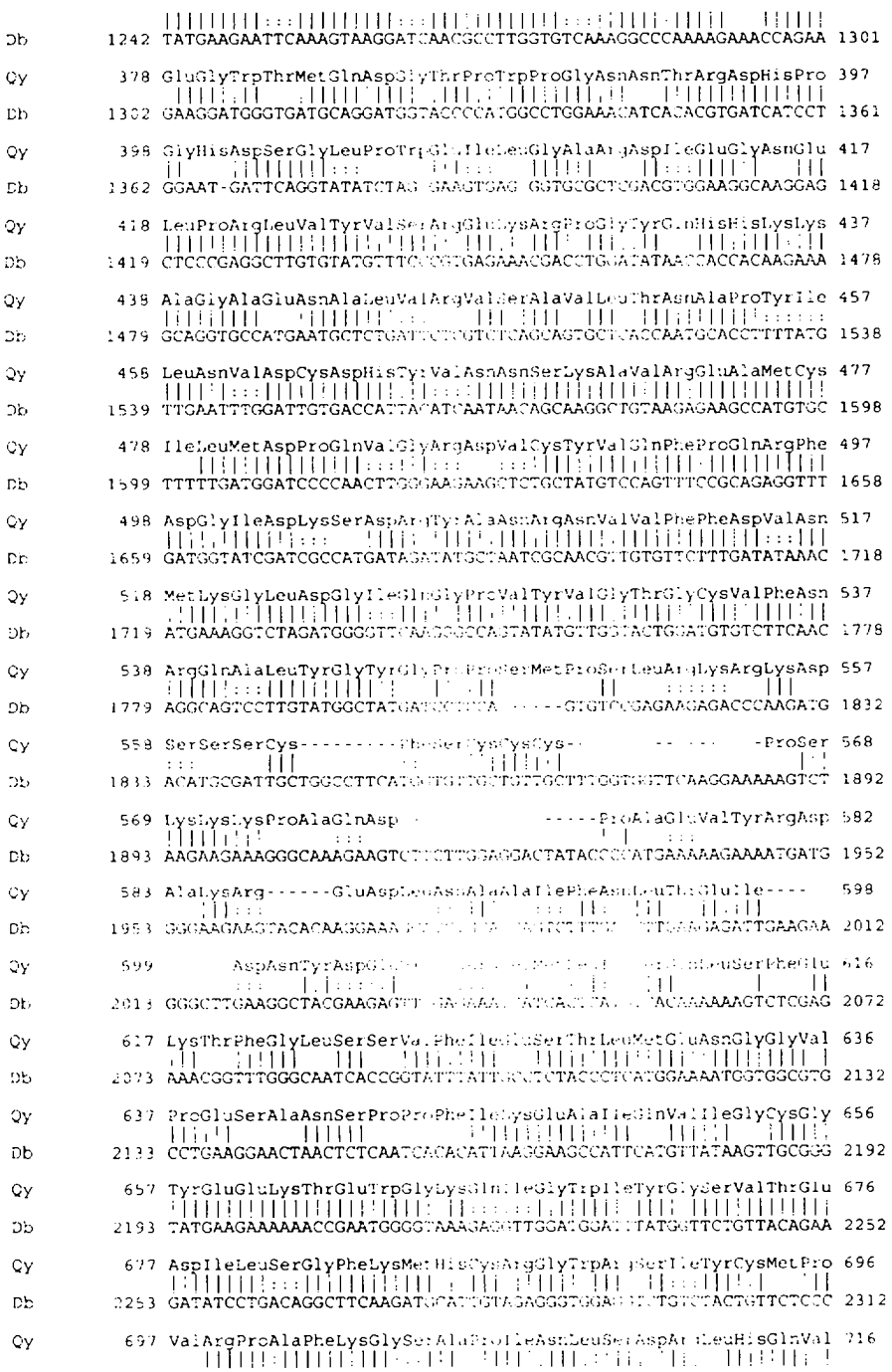
Qy       278 LeuAlaValAspTyrProValAspLysValSerCysTyrValSerAspAspGlyAlaAla 297
             '|:::|||.||||.|||||||'|||  |||| .|||||.|||||||.||||.:::
Db      1002 CTATCCGTTGATTATCCTGTCGACAAGGTCAGTTGTTATGTATCAGATGATGGTGCATCC 1061

Qy       298 MetLeuSerPheGluSerLeuValGluThrAlaGluPheAlaArgLysTrpValProPhe 317
             |||||   ||.:::|.||||   |: |||:|||||||||||||||||:||||'||||
Db      1062 ATGCTCCTTTTCGACTCCCTGGCAGAAACTGCTGAGTTTGCTAGAAAGTGGGTTCCATTT 1121

Qy       318 CysLysLysPheSerIleGluProArgAlaProGluPheTyrPheSerGlnLysIleAsp 337
             ||||||||   :::|||||||||||'||||.|||||||||||||:::||||||||.||
Db      1122 TGCAAGAAGCATAACATTGAGCCAAGGGCTCCTGAGTTCTACTTCACTCAGAAGATTGAC 1181

Qy       338 TyrLeuLysAspLysValGlnProSerPheValLysGluArgArgAlaMetLysArgAsp 357
             |||||||||||||||||   |||:::|||'||||||||||||||||||||||||||| :::
Db      1182 TACTTGAAAGACAAAGTGCATCCCAACTTTGTGAAGGAGCGCAGAGCTATGAAAAGAGAA 1241

Qy       358 TyrGluGluTyrLysValArgValAsnAlaLeuValAlaLysAlaGlnLysThrProGlu 377
```



```
Db       2313 AAGAGACCAGCTTTTAAGGGATCTGCTCCCATTAATCTATCAGATAGGTTGCACCAAGTT 2372

Qy        717 LeuArgTrpAlaLeuGlySerValGluIlePhePheSerArgHisCysProLeuTrpTyr 736
              ||||||||||||||||||||||||:::|||||||||   |||   ||||||||||||||||
Db       2373 CTGCGATGGGCACTAGGCTCTATTGAGATTTTCCTTAGTCATCACTGTCCTCTATGGTAT 2432

Qy        737 GlyPheGlyGlyGlyArgLeuLysTrpLeuGlnArgLeuAlaTyrIleAsnThrIleVal 756
              |||:::   ||||||:::||||||   |||:::||||||||||||||||||||||||||||
Db       2433 GGCTAT---GGGGGAAAGTTGAAGTTGCTGGAGAGGCTTGCTTACATCAACACCATCGTT 2489

Qy        757 TyrProPheThrSerLeuProLeuIleAlaTyrCysThrIleProAlaValCysLeuLeu 776
              |||||||||||||:::|||||:::||||||||||   ||||||||||||||||||||||
Db       2490 TACCCTTTCACCTCCATTCCCTTACTTGCCTACTGTACTACTCCAGCAGTCTGCCTTCTG 2549

Qy        777 ThrGlyLysPheIleIleProThrLeuSerAsnLeuAlaSerMetLeu-PheLeuGlyLe 796
              |||||||||||||||||||||||||:::|||||||||||||:::    ||||||   ||
Db       2550 ACAGGAAAATTTATCATTCCTACTCTGAACAACCTTGCTAGCATATGGTTTCCTGGCCCT 2609

Qy        796 uPheIleSerIleIleValThrAlaValLeuGluLeuArgTrpSerGlyValSerIleGl 816
              ||||||||||||||||   |||:::|||   ||||||   ||||||||||||||||||::
Db       2610 TTTCATCTCAATCATAGCAACATCTGTGTGGGAGTTGGA-TGGAGTGGAGTCAGCATCCA 2668

Qy        816 uAspLeuTrpArgAsnGluGlnPheTrpValIleGlyGlyValSerAlaHisLeuPheAl 836
              :||||||   ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2669 GGACTTGGGGCGTAATGAGCAATTTTGGGTTATCGGTGGTGTGTCAGGTCATCTTTTTGC 2728

Qy        836 aValPheGlnGlyPheLeuLysMetLeuAlaGlyIleAspThrAsnPheThrValThrAl 856
              ||||||||||||||||   |||:::|||   ||:::||||||||||||||||||||||::
Db       2729 TGTCTTCCAAGGCTTCTTCAAGGTCCTTGGAGGAGTTGACACTAACTTCACTGTTACATC 2788

Qy        856 aLysAlaAlaGluAspAlaGluPheGlyGluLeuTyrMetValLysTrpThrThrLeuLe 876
              :|||:::|||:::|||||||||||||||||||||||||:::   ||||||||||||||||
Db       2789 AAAATCAGCAGACGATGCCGAGTTTGGAGAGCTGTACCTCTTCAAATGGACCACCCTCCT 2848

Qy        876 uIleProProThrThrLeuLeuIleIleAsnMetSerGly---CysAlaGlyPheSerAs 895
              ||||||||||||||||:::|||:::||||||   |||      ||||||   :||
Db       2849 CATCCCACCAACCACTCTAATCATCTTGAATATGGTTGGAGTTGTAGCAGGAGTATCGGC 2908

Qy        895 pAlaLeuAsnLysGlyTyrGluAlaTrpGlyProLeuPheGlyLysValPhePheAlaPh 915
                 :::|||   ||||||   :::||||||||||||||||||||:::||||||||||||
Db       2909 ACGAATAAACAACGGATATGGATCATGGGGTCCTTTATTTGGGAAGCTATTTTTCGCTTT 2968

Qy        915 eTrpValIleLeuHisLeuTyrProPheLeuLysGlyLeuMetGlyArgGlnAsnLeuTh 935
              ||||||||||:::|||||  ||||||||||||||||||||||||||:::||||||   ||
Db       2969 CTGGGTCATTGTCCATCTCTATCCTTTCCTCAAAGGTCTGATGGGAAAGCAAAACAGGAC 3028

Qy        935 rProThrIleValValLeuTrpSerValLeuLeuAlaSerValPheSerLeuValTrpVa 955
              |||||||||||||||||||||||||||||||||||||||||||:||||||||:::|||||
Db       3029 TCCTACAATTGTTGTCCTCTGGTCTGTACTTCTTGCATCTATTTTCTCATTGATTTGGGT 3088

Qy        955 lLysIleAsnProPheVa:AsnLysValAspAsnThrLeuValAlaGluThrCysIleSe 975
              |:::|:::|||||:::   |||      ::::::   :::
Db       3089 TAGAATTGATCCCTTCTTGCCCAAGCAAACTGGCCCAATTCTCAAACAA------TGTGG 3142

Qy        975 rIleAspCys 978
              ::::::|||
Db       3143 AGTGGAGTGC 3152

RESULT 15
AF200533
LOCUS       AF200533               3795 bp    mRNA    linear   PLN 31-AUG-2000
DEFINITION  Zea mays cellulose synthase-9 (CesA-9) mRNA, complete cds.
ACCESSION   AF200533
VERSION     AF200533.1  GI:9622889
KEYWORDS    .
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 3795)
  AUTHORS   Holland,N., Holland,D., Helentjaris,T., Dhugga,K.S.,
            Xoconostle-Cazares,B. and Delmer,D.P.
```

```
TITLE     A comparative analysis of the plant cellulose synthase (CesA) gene
          family
JOURNAL   Plant Physiol. 123 (4), 1313-1324 (2000)
MEDLINE   20398328
 PUBMED   10938350
FERENCE   2  (bases 1 to 3795)
AUTHORS   Dhugga,K.S. and Helentjaris,T.G.
TITLE     Direct Submission
JOURNAL   Submitted (01-NOV-1999) Agronomic Traits, Pioneer Hi-Bred
          International, Inc., 7250 NW 62nd Avenue, Johnston, IA 50131, USA
ATURES             Location/Qualifiers
   source          1. .3795
                   /organism="Zea mays"
                   /mol_type="mRNA"
                   /db_xref="taxon:4577"
                   /note="identified and isolated from the maize genome
                   database cDNA collection of Pioneer Hi Bred International,
                   Inc"
   gene            1. .3795
                   /gene="CesA-9"
   CDS             238. .3477
                   /gene="CesA-9"
                   /note="CESA-9"
                   /codon_start=1
                   /product="cellulose synthase-9"
                   /protein_id="AAF89969.1"
                   /db_xref="GI:9622890"
                   /translation="MEGDADGVKSGRRGGGQVCQICGDGVGTTAEGDVFTACDVCGFP
                   VCRPCYEYERKDGTQACPQCKNKYKRHKGSPAIRGEEGDDTDADDASDFNYPASGNDD
                   QKQKIADRMRSWRMNAGGSGDVGRPKYDSGEIGLTKYDSGEIPRGYIPSVTNSQISGE
                   IPGASPDHHMMSPTGNIGRRAPFPYMNHSSNPSREFSGSVGNVAWKERVDGWKMKQDK
                   GTIPMTNGTSIAPSEGRGVGDIDASTDYNMEDALLNDETRQPLSRKVPLPSSRINPYR
                   MVIVLRLIVLSIFLHYRITNPVRNAYPLWLLSVICEIWFALSWILDQFPKWFPINRET
                   YLDRLALRYDREGEPSQLAAVDIFVSTVDPMKEPPLVTANTVLSILAVDYPVDKVSCY
                   VSDDGAAMLTFDALAETSEFARKWVPFVKKYNIEPRAPEWYFSQKIDYLKDKVHPSFV
                   KDRRAMKREYEEFKIRVNGLVAKAQKVPEEGWIMQDGTPWPGNNTRDHPGMIQVFLGH
                   SGGLDTEGNELPRLVYVSREKRPGFQHHKKAGAMNALVRVSAVLTNGQYMLNLDCDHY
                   INNSKALREAMCFLMDPNLGRSVCYVQFPQRFDGIDRNDRYANRNTVFFDINLRGLDG
                   IQGPVYVGTGCVFNRTALYGYEPPIKQKKGGFLSSLCGGRKKGSKSKKGSDKKKSQKH
                   VDSSVPVFNLEDIEEGVEGAGFDDEKSLLMSQMSLEKRFGQSAAFVASTLMEYGGVPQ
                   SATPESLLKEAIHVISCGYEDKIEWGTEIGWIYGSVTEDILTGFKMHARGWRSIYCMP
                   KRPAFKGSAPINLSDRLNQVLRWALGSVEILFSRHCPLWYGYGGRLKFLERFAYINTT
                   IYPLTSLPLLIYCILPAICLLTGKFIIPEISNFASIWFISLFISIFATGILEMRWSGV
                   GIDEWWRNEQFWVIGGISAHLFAVFQGLLKVLAGIDTNFTVTSKASDEDGDFAELYMF
                   KWTTLLIPPTTILIINLVGVVAGISYAINSGYQSWGPLFGKLFFAFWVIVHLYPFLKG
                   LMGRQNRTPTIVVVWAILLASIFSLLWVRIDPFTNRVTGPDTRTCGINC"
ASE COUNT      925 a    927 c    995 g    944 t      4 others
RIGIN

lignment Scores:
red. No.:              6.33e-285      Length:        3795
core:                  3366.50        Matches:       648
ercent Similarity:     72.58%         Conservative:  133
est Local Similarity:  60.22%         Mismatches:    183
uery Match:            64.25%         Indels:        114
B:                     8              Gaps:          17

S-09-980-043-2 (1-978) x AF200533 (1-3795)

y          5 GlyAlaProIleCysHisThrCysGlyGluGlnValGlyHisAspAlaAsnGlyGluLeu 24
b        280 GGAGGGCAGGTGTGCCAGATCTGCGGCGATGGCGTGGGCACTACGGCGGAGGGAGACGTC 339

y         25 PheValAlaCysHisGluCysSerTyrProMetCysLysSerCysPheGluPheGluIle 44
b        340 TTCACCGCCTGCGACGTCTGCGGGTTCCCGGTGTGCCGCCCCTGCTACGAGTACGAGCGC 399

y         45 AsnGluGlyArgLysValCysLeuArgCys----------------------GlySer 56
b        400 AAGGACGGCACACAAGCGTGCCCCCAGTGCAAAAACAAGTACAAGCGCCACAAGGGGAGT 459

y         57 Pro------------------------------------------------------ 57
```

```
Db       460 CCAGCGATGCGAGTTCAAGGAACGAATATGTATGCGATGCCGATGATGCTAGCGACTTCAAC 519

Qy        58 TyrAspGluAsnGluLeuAspAspAlaLysLysGlySerGlyAsnGlnSerThrMet 77

Db       520 TACCCTGCATCTGGTAATGACGACGAGAAGCAGAAGATTGCTGACAGGATGCGCAGCTGG 579

Qy        78 AlaSerHisGluAspSerGlnAspGlyLys---------- 88

Db       580 CGCATGAATGCTGGCAGCGGCATGTTTGCCCCCAAGTATGACAGTGGTGAGATC 639

Qy        89 --------------IleHisAlaArgHisIleSerSerVal 97

Db       640 GGGCTTACTAAGTACAGTGCTATGCTGGGGATACATCCCGTCAGTCACTAAC 699

Qy        97 ------------------------------------------ 97

Db       700 AGCCAGATTGACTTCTCCTGACCATCATATCATGTCTCCTACT 759

Qy        98 ----------------------------SerThrValAspSer 102

Db       760 GGGAACATTGGCATGCTGGCCGATATCGCTGTAGGAATCATTCATCAAATCTGTCGAGG 819

Qy       103 GluMetAsnAspGluTyrGlyAsnProLeuTrpLysAsnArgValLysSerCysLysAsp 122

Db       820 GAATTCTCTGGTATGTTGGGAATGGTTGTGAAAGAGAGGGTTGATGGCTGGAAAATG 879

Qy       123 LysGluAsnLys----LysGlyArgSerProLysAlaGluThrGluPro 138

Db       880 AAGCAGGACAAGGTAACAATTCGCATGACGAAGGCACAAGCATTGCTCCCTCTGAGGGC 939

Qy       139 AlaGlnVal----ProGlnGlnGlnMetGluGluLys---ProSer 152

Db       940 CGGGGTGTTGGTGATTATTGATGCATCAACTGATTACAACATGGAAGATGCCTTATTAAAC 999

Qy       153 AlaGluAlaSerGlyProLeuSerIleValTyrProIleProArgAsnLysLeuThrPro 172

Db      1000 GATGAAACTCGCCAGCTCTATCTAGGAAAGTTGCACTTCCTTCCTCCAGGATAAATCCA 1059

Qy       173 TyrArgAlaValIleIleMetArgLeuValIleLeuGlyLeuPhePheHisPheArgIle 192

Db      1060 TACAGGATGGTCATTGTGCTACGATTGATTGTTCTAAGCATCTTCTTGCACTACCGGATC 1119

Qy       193 ThrArgAsnProValAspSerAlaPheGlyLeuTrpLeuThrSerValIleCysGluIleTrp 212

Db      1120 ACAAATCCTGTGCGTAATGCATACGGACTGTGGTTCTATCTGTTATATGTGAGATCTGG 1179

Qy       213 PheAlaPheSerTrpValLeuAspGlnPheProLysTrpAsnProValAsnArgGluThr 232

Db      1180 TTTGCTCTTTCCTGGATATTGGACCAGTTTCCAAAGTGGTTTCCAATCAACCGCGAGACT 1239

Qy       233 TyrIleGluArgLeuSerAlaArgTyrGluArgGluGlyGluProSerGlnLeuAlaGly 252

Db      1240 TACATTGAGAGACTGGCATTAAGATATGAGCGAAGGTGAGCCATCTCAGTTGGCTGCT 1299

Qy       253 ValAspPhePheValSerThrValAspProLeuLysGluProProLeuIleThrAlaAsn 272

Db      1300 GTTGACATTTTTGTCAGTACTGTGGACCCAATGAAGGAGCCTCCTCTTGTCACTGCCAAT 1359

Qy       273 ThrValLeuSerIleLeuAlaValAspTyrProValAspLysValSerCysTyrValSer 292

Db      1360 ACCGTGCTATCCATTCTCGCTGTGGATTATCTTGTGGATAAGGTCTCTTGCTATGTATCT 1419

Qy       293 AspAspGlyAlaAlaMetLeuSerPheGluSerLeuValGluThrAlaGluPheAlaArg 312

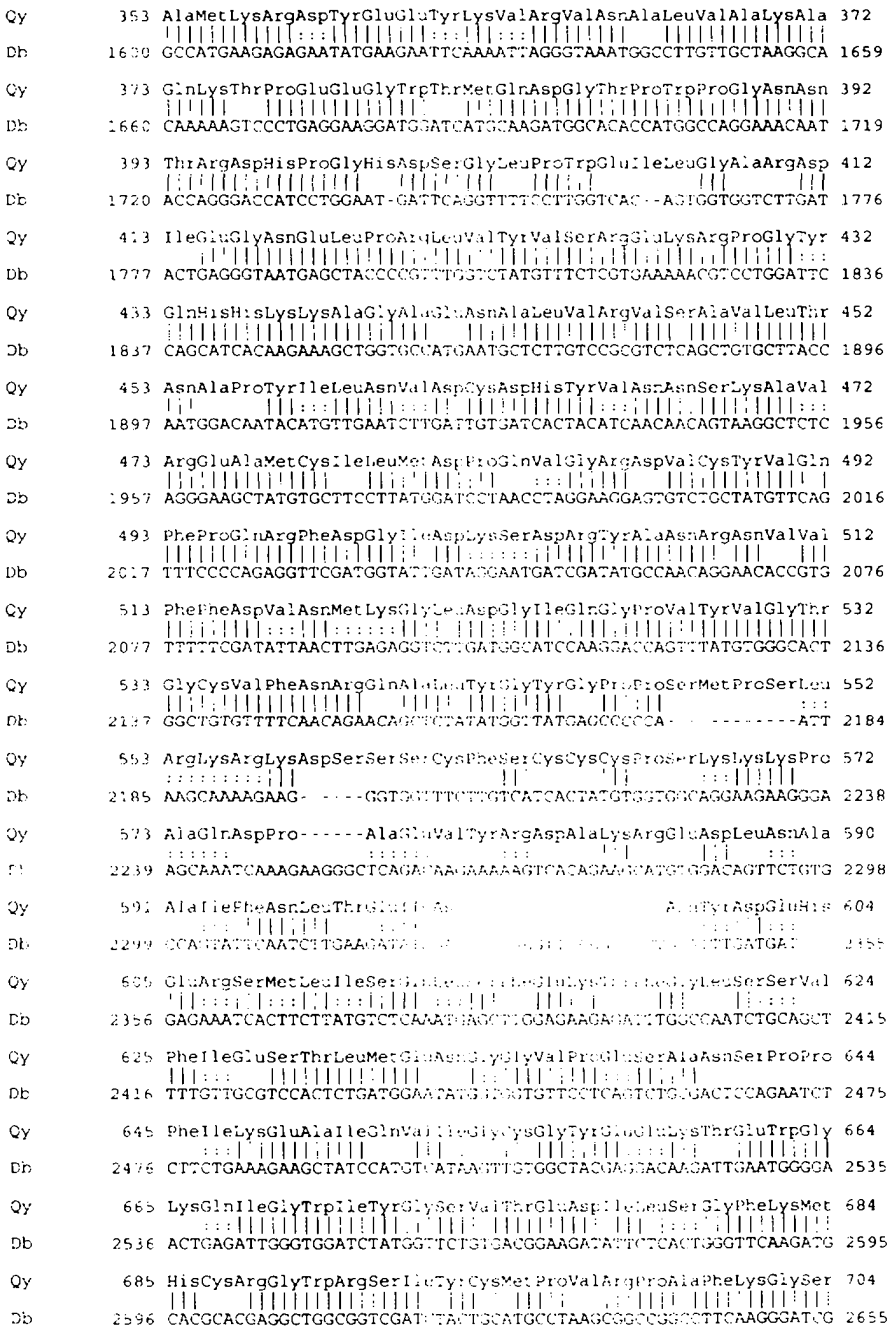
Db      1420 GATGATGGAGCTGCTATGCTGACATTTGATGCACTAGCTGAGACTTCAGAGTTTGCTAGA 1479

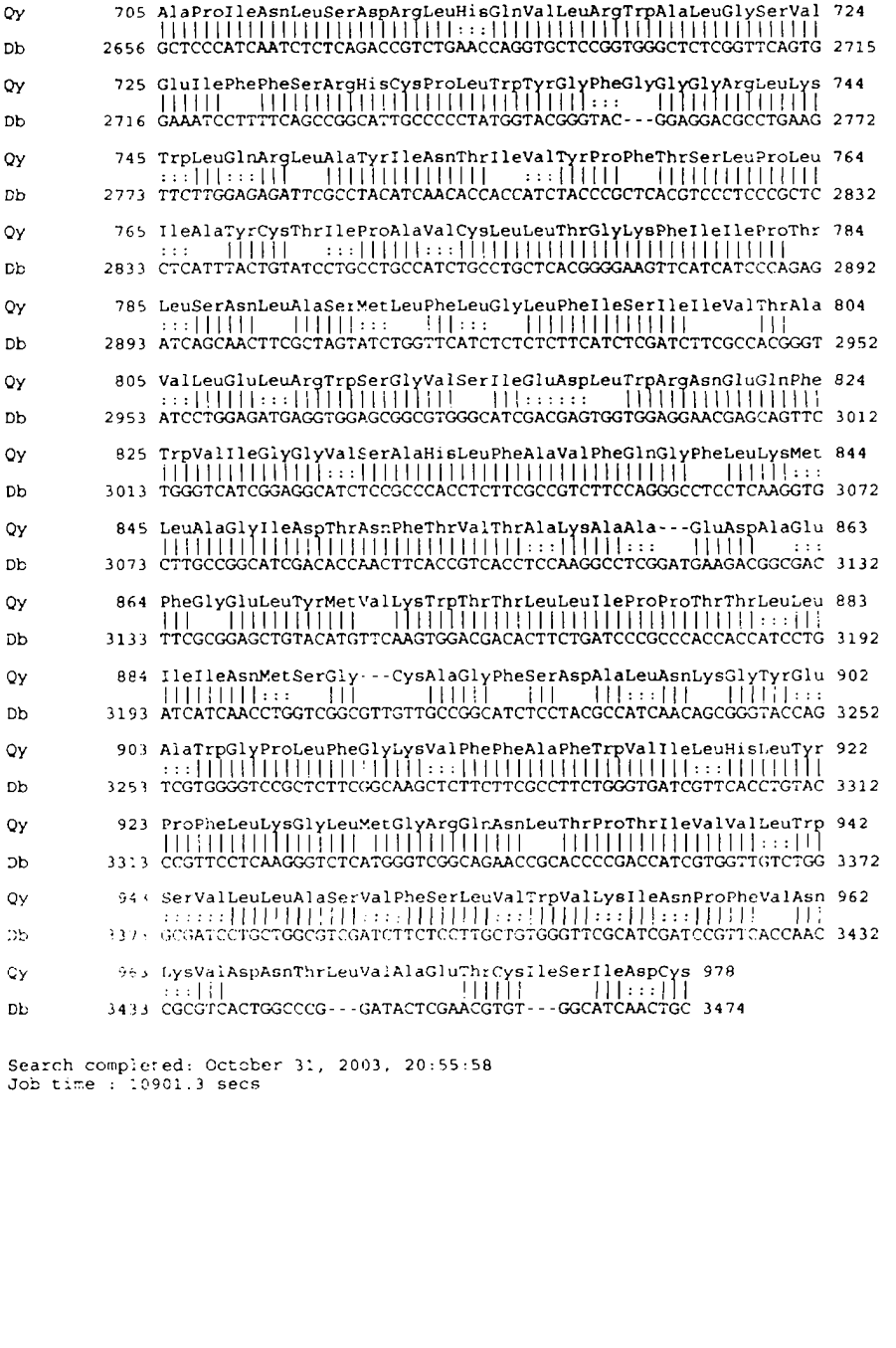
Qy       313 LysTrpValProPheCysLysLysPheSerIleGluProArgAlaProGluPheTyrPhe 332

Db      1480 AAATGGGTGCCATTTGTTAAGAAGTACAACATTGAACCTAGAGCTCCTGAATGGTACTTC 1539

Qy       333 SerGlnLysIleAspTyrLeuLysAspLysValGlnProSerPheValLysGluArgArg 352

Db      1540 TCCCAGAAAATTGATTACTTGAAGGACAAAGTGCACCCTTCATTTGTTAAAGACCGCCGG 1599
```





Search completed: October 31, 2003, 20:55:58
Job time : 10901.3 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

CM nucleic - nucleic search, using sw model

Run on: October 30, 2003, 05:33:15 ; Search time 11515.9 Seconds
(without alignments)
11481.594 Million cell updates/sec

Title: US-09-980-043-1
Perfect score: 3232
Sequence: 1 gtcgacccacgcgtccgtc...aagaaaaaaaaaaaaaaaaa 3232

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481486 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3232	100.0	3232	8	AF072131	AF072131 Populus t
2	1989.8	61.6	3186	8	GHU58283	U58283 Gossypium h
3	1989.8	61.6	3328	6	AR164434	AR164434 Sequence
4	1677	51.9	3109	8	AY099636	AY099636 Arabidops
5	1666.8	51.6	3145	8	AF323039	AF323039 Zinnia el
6	1666	51.5	2958	8	AF267742	AF267742 Arabidops
7	1162.2	36.0	3723	8	AF150630	AF150630 Gossypium
8	1148.8	35.5	3752	8	AF200525	AF200525 Zea mays
9	1146.2	35.5	3222	6	AX652952	AX652952 Sequence
10	1137.8	35.2	3676	8	AF200529	AF200529 Zea mays
11	1137.6	35.2	3725	8	AF200526	AF200526 Zea mays
12	1135.4	35.1	3229	8	BT002335	BT002335 Arabidops
13	1133.8	35.1	3614	6	AR267559	AR267559 Sequence
14	1133.8	35.1	3614	6	AX030946	AX030946 Sequence
15	1133.8	35.1	3614	6	BD022678	BD022678 Manipulat
16	1132.4	35.0	3682	8	AF027174	AF027174 Arabidops
17	1128.6	34.9	3795	8	AF200533	AF200533 Zea mays
18	1128.6	34.9	3799	6	AX338680	AX338680 Sequence
19	1106.8	34.2	3081	8	AF088917	AF088917 Arabidops
20	1106.8	34.2	3081	8	BT004543	BT004543 Arabidops
21	1106.8	34.2	3355	8	AY139754	AY139754 Arabidops
22	1106.2	34.2	3968	8	AF200531	AF200531 Zea mays
23	1103.2	34.1	2830	6	AX338676	AX338676 Sequence
24	1102.8	34.1	2828	8	AF200527	AF200527 Zea mays
25	1102.8	34.1	3745	8	AF200528	AF200528 Zea mays
26	1097.8	34.0	3603	6	AR267557	AR267557 Sequence
27	1097.8	34.0	3603	6	AX030942	AX030942 Sequence
28	1097.8	34.0	3603	6	BD022676	BD022676 Manipulat
29	1097.8	34.0	3763	8	BT008654	BT008654 Arabidops
30	1097	33.9	3538	8	AF200530	AF200530 Zea mays
31	1096.2	33.9	3673	6	AR267560	AR267560 Sequence
32	1096.2	33.9	3673	6	AX030948	AX030948 Sequence
33	1096.2	33.9	3673	6	BD022679	BD022679 Manipulat
34	1091.6	33.8	3395	8	AF081534	AF081534 Populus x
35	1091.4	33.8	3264	6	AX653232	AX653232 Sequence
36	1084.8	33.6	3181	8	BT006111	BT006111 Arabidops
37	1084.8	33.6	3377	8	BT005710	BT005710 Arabidops
38	1082	33.5	3812	8	AF200532	AF200532 Zea mays
39	1070.2	33.1	3847	8	AF304374	AF304374 Nicotiana
40	1069.6	32.8	3277	8	AY095297	AY095297 Populus t
41	1056.4	32.7	3168	8	AF458083	AF458083 Arabidops
42	1051	32.5	3255	6	AX505864	AX505864 Sequence
43	1051	32.5	3771	8	AF027173	AF027173 Arabidops
44	1048.6	32.4	3255	6	AX507835	AX507835 Sequence
45	1043	32.3	3828	6	AR267558	AR267558 Sequence

ALIGNMENTS

RESULT 1
AF072131
LOCUS AF072131 3232 bp mRNA linear PLN 26-APR-2001
DEFINITION Populus tremuloides secondary xylem cellulose synthase (celA) mRNA, complete cds.
ACCESSION AF072131
VERSION AF072131.1 GI:4115904
KEYWORDS .
SOURCE Populus tremuloides (quaking aspen)
ORGANISM Populus tremuloides
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids I; Malpighiales; Salicaceae; Populus.
REFERENCE 1 (bases 1 to 3232)
AUTHORS Wu,L., Joshi,C.P. and Chiang,V.L.

```
 TITLE     A xylem-specific cellulose synthase gene from aspen (Populus
           tremuloides) is responsive to mechanical stress
 JOURNAL   Plant J. 22 (6), 495-502 (2000)
 MEDLINE   20344632
  PUBMED   10886769
CFERENCE   2  (bases 1 to 3232)
 AUTHORS   Wu,L., Joshi,C.P. and Chiang,V.L.
 TITLE     Direct Submission
 JOURNAL   Submitted (14-JUN-1998) IWR/Forestry, Michigan Tech University,
           1400, Townsend Drive, Houghton, MI 49931, USA
CFERENCE   3  (bases 1 to 3232)
 AUTHORS   Wu,L., Joshi,C.P. and Chiang,V.L.
 TITLE     Direct Submission
 JOURNAL   Submitted (07-JAN-1999) IWR/Forestry, Michigan Tech University,
           1400, Townsend Drive, Houghton, MI 49931, USA
 REMARK    Sequence update by submitter
OMMENT     On Jan 8, 1999 this sequence version replaced gi:5243277.
EATURES             Location/Qualifiers
     source         1. .3232
                    /organism="Populus tremuloides"
                    /mol_type="mRNA"
                    /db_xref="taxon:3693"
                    /tissue_type="secondary xylem"
     gene           1. .3232
                    /gene="celA"
     CDS            69. .3005
                    /gene="celA"
                    /codon_start=1
                    /product="secondary xylem cellulose synthase"
                    /protein_id="AAD03417.1"
                    /db_xref="GI:4115905"
                    /translation="MMESGAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEI
                    NEGRKVCLRCGSPYDENLLDDVEKKGSGNQSTMASHLNDSQDVGIHARHISSVSTVDS
                    EMNDEYGNPIWKNRVKSCKDKENKKKKRSPKAETEPAQVPTEQQMEEKPSAEASEPLS
                    IVYPIPRNKLTPYRAVIIMRLVILGLFFHFRITNPVDSAFGLWLTSVICEIWFAFSWV
                    LDQFPKWNPVNRETYIERLSARYEREGEPSQLAGVDFFVSTVDPLKEPPLITANTVLS
                    ILAVDYPVDKVSCYVSDDGAAMLSFESLVETAEFARKWVPFCKKFSIEPRAPEFYFSQ
                    KIDYLKDKVQPSFVKERRAMKRDYEEYKVRVNALVAKAQKTPEEGWTMQDGTPWPGNN
                    TRDHPGHDSGLPWEILGARDIEGNELPRLVYVSREKRPGYQHHKKAGAENALVRVSAV
                    LTNAPYILNVDCDHYVNNSKAVREAMCILMDPQVGRDVCYVQFPQRFDGIDKSDRYAN
                    RNVVFFDVNMKGLDGIQGPVYVGTGCVFNRQALYGYGPPSMPSLRKRKDSSSCFSCCC
                    PSKKKPAQDPAEVYRDAKREDLNAAIFNLTEIDNYDEHERSMLISQLSFEKTFGLSSV
                    FIESTLMENGGVPESANSPPFIKEAIQVIGCGYEEKTEWGKQIGWIYGSVTEDILSGF
                    KMHCRGWRSIYCMPVRPAFKGSAPINLSDRLHQVLRWALGSVEIFFSRHCPLWYGFGG
                    GRLKWLQRLAYINTIVYPFTSLPLIAYCTIPAVCLLTGKFIIPTLSNLASMLFLGLFI
                    SIIVTAVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQGFLKMLAGIDTNFTVTA
                    KAAEDAEFGELYMVKWTTLLIPPTTLLIINMSGCAGFSDALNKGYEAWGPLFGKVFFA
                    FWVILHLYPFLKGLMGRQNLTPTIVVLWSVLLASVFSLVWVKINPFVNKVDNTLVAET
                    CISIDC"
ASE COUNT      895 a    668 c    772 g    897 t
RIGIN

 Query Match           100.0%;  Score 3232;  DB 8;  Length 3232;
 Best Local Similarity  100.0%;  Pred. No. 0;
 Matches 3232;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTCGACCCACGCGTCCGTCTTGAAAGAATATGAAGTTGTAAAGAGCTGGTAAAGTGGTAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACCCACGCGTCCGTCTTGAAAGAATATGAAGTTGTAAAGAGCTGGTAAAGTGGTAA 60

Qy         61 TAAGCAAGATGATGGAATCTGGGGCTCCTATATGCCATACCTGTGGTGAACAGGTGGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TAAGCAAGATGATGGAATCTGGGGCTCCTATATGCCATACCTGTGGTGAACAGGTGGGGC 120

Qy        121 ATGATGCAAATGGGGAGCTATTTGTGGCTTGCCATGAGTGTAGCTATCCCATGTGCAAGT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ATGATGCAAATGGGGAGCTATTTGTGGCTTGCCATGAGTGTAGCTATCCCATGTGCAAGT 180

Qy        181 CTTGTTTCGAGTTTGAAATCAATGAGGGCCGGAAAGTTTGCTTGCGGTGTGGCTCGCCAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CTTGTTTCGAGTTTGAAATCAATGAGGGCCGGAAAGTTTGCTTGCGGTGTGGCTCGCCAT 240

Qy        241 ATGATGAGAACTTGCTGGATGATGTAGAAAAGAAGGGGTCTGGCAATCAATCCACAATGG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ATGATGAGAACTTGCTGGATGATGTAGAAAAGAAGGGGTCTGGCAATCAATCCACAATGG 300

Qy        301 CATCTCAGCTTCAATTCATCTCAGTATTCTTAATTCATGCTAGACATATCAGTAGTGTGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CATCTCAGCTTCAATTCATCTCAGTATTCTTAATTCATGCTAGACATATCAGTAGTGTGT 360

Qy        361 CCACTGTGGAATAGTAAAAGAAGAATGATGAAGAAGTCAATCCAATTTGGAAGAATCGGGTGA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CCACTGTGGAATAGTAAAAGAAGAATGATGAAGAAGTCAATCCAATTTGGAAGAATCGGGTGA 420

Qy        421 AGAGCTGTAAGGATAAAAAGAGAACTAAAAAAAAAAAAGAAGTCCTAAGGCTGAAACTGAAC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGAGCTGTAAGGATAAAAAGAGAACTAAAAAAAAAAAAGAAGTCCTAAGGCTGAAACTGAAC 480

Qy        481 CAGCTCAAAGAGAAGATGGAAGGTAAAAAGTTCGGCTTCTACGTCATGAAACCGTCTGCAGAGGCTTCGGAGC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CAGCTCAAAGAGAAGATGGAAGGTAAAAAGTTCGGCTTCTACGTCATGAAACCGTCTGCAGAGGCTTCGGAGC 540

Qy        541 CGCTTTCAATTGTTTATGCAATCTTAGCAGCAGCTCACACCATACAGAGCAGTGATCA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CGCTTTCAATTGTTTATGCAATCTTAGCAGCAGCTCACACCATACAGAGCAGTGATCA 600

Qy        601 TTATGCGATTGGTCAATCTTGGGCTTCTTTTCTTCAGTTCAGAATAACAAATCCTGTCGATA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TTATGCGATTGGTCAATCTTGGGCTTCTTTTCTTCAGTTCAGAATAACAAATCCTGTCGATA 660

Qy        661 GTGCCTTTGGGCCTGTGGTTTACTTCTTGGATATGTGGAGATCTGGTTTGCATTTTCTTGGG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GTGCCTTTGGGCCTGTGGTTTACTTCTTGGATATGTGGAGATCTGGTTTGCATTTTCTTGGG 720

Qy        721 TGTTGGATCAGTTTTTCAAGTGGAATCCTCTTAATAGAGAAACGTATATCGAAAGGCTGT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TGTTGGATCAGTTTTTCAAGTGGAATCCTCTTAATAGAGAAACGTATATCGAAAGGCTGT 780

Qy        781 CGGCAAGGTATGAAAGAGAGGGTGAGCCTTCTTAGCTTGCTGGTGTGGATTTTTTCGTGA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CGGCAAGGTATGAAAGAGAGGGTGAGCCTTCTTAGCTTGCTGGTGTGGATTTTTTCGTGA 840

Qy        841 GTACTGTTGATCCGTTGAAGGAACCGGCCATTGATCACTGCCAATACAGTCCTTTCCATCC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GTACTGTTGATCCGTTGAAGGAACCGGCCATTGATCACTGCCAATACAGTCCTTTCCATCC 900

Qy        901 TTGCTGTGGACTATTCCGTCGATAAAGTCTCTTTGCTACGTGTCTGATGATGGTGCAGCTA 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTGCTGTGGACTATTCCGTCGATAAAGTCTCTTTGCTACGTGTCTGATGATGGTGCAGCTA 960

Qy        961 TGCTTTCATTTGAATCTCTTGTAGAAACAGCTCGAGTTTGCAAGGAAGTGGGTTCCGTTCT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 TGCTTTCATTTGAATCTCTTGTAGAAACAGCTCGAGTTTGCAAGGAAGTGGGTTCCGTTCT 1020

Qy       1021 GCAAAAAATTCTCAATTGAACCAAGAGCTACTCGAGTTTTACTTCTCACAGAAAATTGATT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GCAAAAAATTCTCAATTGAACCAAGAGCTACTCGAGTTTTACTTCTCACAGAAAATTGATT 1080

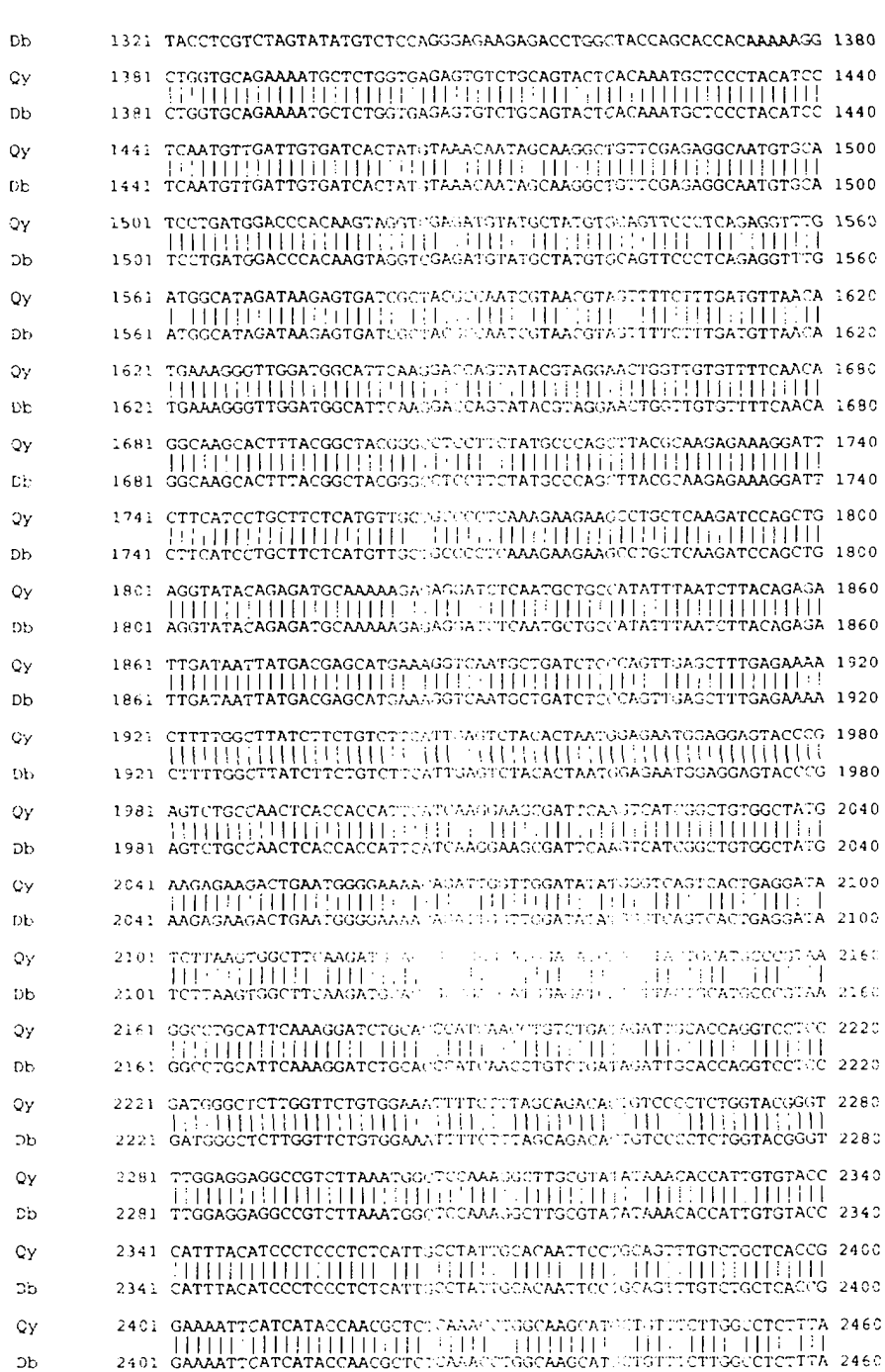
Qy       1081 ACTTGAAAGACAAGTTTCAACCTTCTTTTCGTGAAAGAACGTAGAGCAATGAAAAGGGATT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 ACTTGAAAGACAAGTTTCAACCTTCTTTTCGTGAAAGAACGTAGAGCAATGAAAAGGGATT 1140

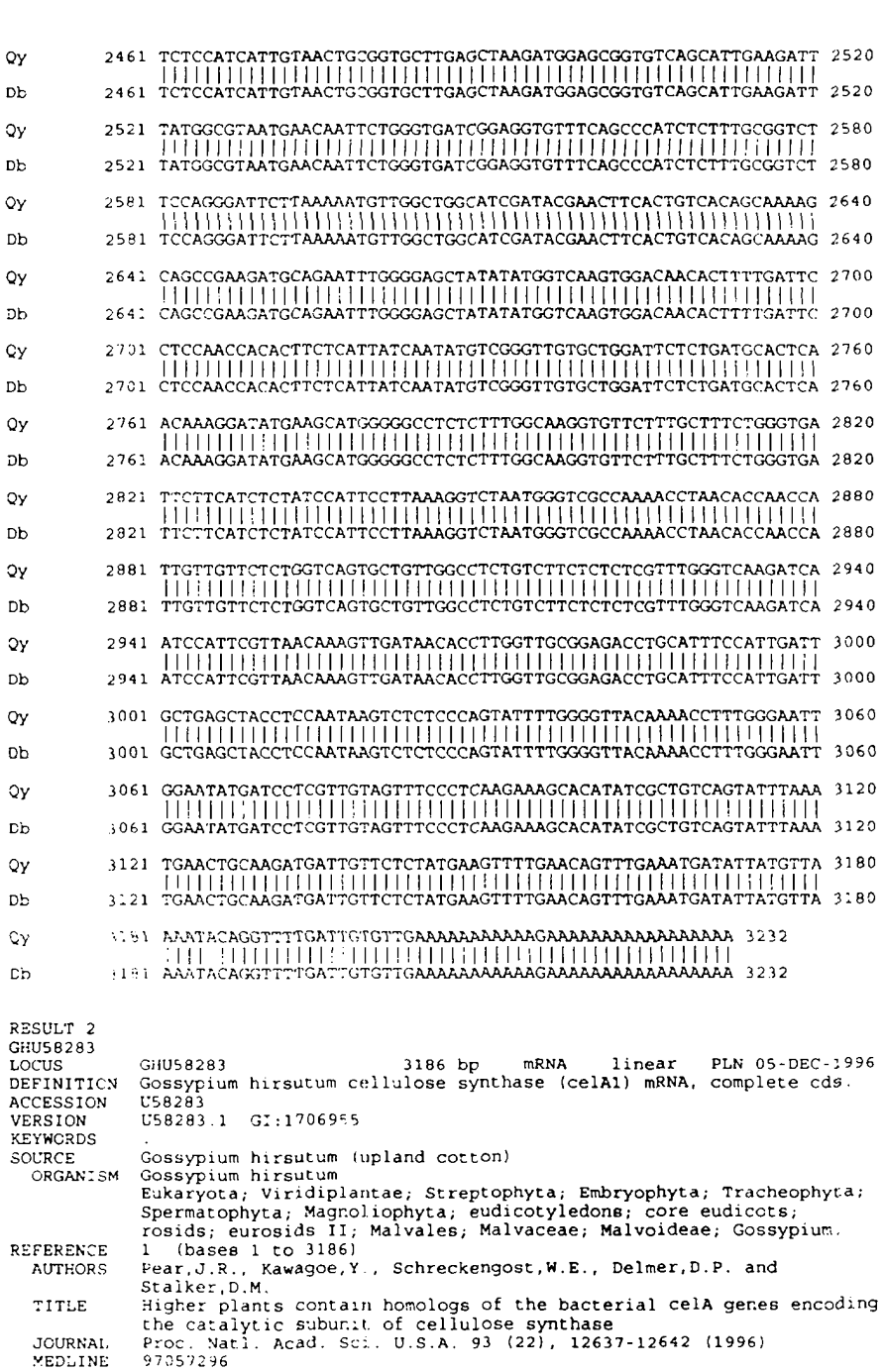
Qy       1141 ATGAAGAGTACAAAGTCCGAGTTAATAGCTCTTGTAGCAAAGGCTCAGAAAACACCTGAAG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ATGAAGAGTACAAAGTCCGAGTTAATAGCTCTTGTAGCAAAGGCTCAGAAAACACCTGAAG 1200

Qy       1201 AAGGATGGACTATGCAAGATGGAACACCTTGGCCTGGGAATAACACACGTGATCACCCTG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 AAGGATGGACTATGCAAGATGGAACACCTTGGCCTGGGAATAACACACGTGATCACCCTG 1260

Qy       1261 GGCATGATTCAGGTCTTCCTTGGGAAATACTGGGAGCTCGTGACATTGAAGGAAATGAAC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 GGCATGATTCAGGTCTTCCTTGGGAAATACTGGGAGCTCGTGACATTGAAGGAAATGAAC 1320

Qy       1321 TACCTCGTCTAGTATATGTCTCCAGGGAGAAGAGACCTGGCTACCAGCACCACAAAAAGG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```





```
RESULT 2
GHU58283
LOCUS       GHU58283                3186 bp    mRNA    linear   PLN 05-DEC-1996
DEFINITION  Gossypium hirsutum cellulose synthase (celA1) mRNA, complete cds.
ACCESSION   U58283
VERSION     U58283.1  GI:1706955
KEYWORDS    .
SOURCE      Gossypium hirsutum (upland cotton)
  ORGANISM  Gossypium hirsutum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium.
REFERENCE   1  (bases 1 to 3186)
  AUTHORS   Pear,J.R., Kawagoe,Y., Schreckengost,W.E., Delmer,D.P. and
            Stalker,D.M.
  TITLE     Higher plants contain homologs of the bacterial celA genes encoding
            the catalytic subunit of cellulose synthase
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 93 (22), 12637-12642 (1996)
  MEDLINE   97057296
```

```
 PUBMED   8901635
FERENCE   2  (bases 1 to 3186)
AUTHORS   Pear,J.R., Kawagoe,Y., Schreckengost,W.E., Delmer,D.P. and
          Stalker,D.M.
TITLE     Direct Submission
JOURNAL   Submitted (14-MAY-1996) J.R. Pear, Calgene, Inc, 1920 5th Street,
          Davis, CA 95616, USA
ATURES             Location/Qualifiers
   source          1. .3186
                   /organism="Gossypium hirsutum"
                   /mol_type="mRNA"
                   /strain="Acala SJ-2"
                   /db_xref="taxon:3635"
                   /tissue_type="fiber"
                   /dev_stage="17-21 days post-anthesis"
   gene            1. .3186
                   /gene="celA1"
   CDS             77. .3001
                   /gene="celA1"
                   /codon_start=1
                   /product="cellulose synthase"
                   /protein_id="AAB37766.1"
                   /db_xref="GI:1706956"
                   /translation="MMESGVPVCHTCGEHVGLNVNGEPFVACHECNFPICKSCFEYDL
                   KEGRKACLRCGSPYDENLLDDVEKATGDQSTMAAHLNKSQDVGIHARHISSVSTLDSE
                   MAEDNGNSIWKNRVESWKEKKNKKKKPATTKVEREAEIPPEQQMEDKPAPDASQPLST
                   IIPIPKSRLAPYRTVIIMRLIILGLFFHYRVTNPVDSAFGLWLTSVICEIWFAFSWVL
                   DQFPKWYPVNRETYIDRLSARYEREGEPDELAAVDFFVSTVDPLKEPPLITANTVLSI
                   LALDYPVDKVSCYISDDGAAMLTFESLVETADFARKWVPFCKKFSIEPRAPEFYFSQK
                   IDYLKDKVQPSFVKERRAMKRDYEEYKIRINALVAKAQKTPDEGWTMQDGTSWPGNNP
                   RDHPGMIQVFLGYSGARDIEGNELPRLVYVSREKRPGYQHHKKAGAENALVRVSAVLT
                   NAPFILNLDCDHYVNNSKAVREAMCFLMDPQVGRDVCYVQFPQRFDGIDRSDRYANRN
                   TVFFDVNMKGLDGIQGPVYVGTGCVFNRQALYGYGPPSMPSFPKSSSSSCSCCPGKK
                   EPKDPSELYRDAKREELDAAIFNLREIDNYDEYERSMLISQTSFEKTFGLSSVFIEST
                   LMENGGVAESANPSTLIKEAIHVISCGYEEKTAWGKEIGWIYGSVTEDILTGFKMHCR
                   GWRSIYCMPLRPAFKGSAPINLSDRLHQVLRWALGSVEIFLSRHCPLWYGFGGGRLKW
                   LQRLAYINTIVYPFTSLPLIAYCSLPAICLLTGKFIIPTLSNLASVLFLGLFLSIIVT
                   AVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQGFLKMLAGIDTNFTVTAKAADD
                   ADFGELYIVKWTTLLIPPTTLLIVNMVGVVAGFSDALNKGYEAWGPLFGKVFFSFWVI
                   LHLYPFLKGLMGRQNRTPTIVVLWSVLLASVFSLVWVRINPFVSTADSTTVSQSCISI
                   DC"
SE COUNT      864 a    654 c    723 g    945 t
IGIN

Query Match             61.6%;  Score 1989.8;  DB 8;  Length 3186;
Best Local Similarity   80.8%;  Pred. No. 0;
Matches 2399;  Conservative    0;  Mismatches  552;  Indels   18;  Gaps    6;

          40 AAAGAGCTGGTAAAGTGGTAATAAGCAAGATGATGGAATCTGGGGCTCCTATATGCCATA 99
             || |||   |  |||| |  |    |  ||||||||||||||||| ,||| | |||||
          48 AAGGAGAAAGATAAGTACTTTTTTTTGAGAATGATGGAATCTGGGGTTCCTGTTTGCCACA 107

         100 CCTGTGGTGAACAGGTGGGGCATGATGCAAATGGGGAGCTATTTGTGGCTTGCCATGAGT 159
             | ||||||||||| || |||    |||  ||||| || |  ||||||||||||||||| | |
         108 CTTGTGGTGAACATGTTGGGTTGAATGTTAATGGTGAACCTTTTGTGGCTTGCCATGAAT 167

         160 GTAGCTATCCCATGTGCAAGTCTTGTTTCGAGTTTGAAATCAATGAGGGCCGGAAAGTTT 219
             ||| |  || || || |||  |||||| |||| ||| |  | || || || || ||| ||
         168 GTAATTTCCCTATTTGTAAGAGTTGTTTTGAGTATGATCTTAAGGAAGGACGAAAAGCTT 227

         220 GCTTGCGGTGTGGCTCGCCATATGATGAGAACTTGCTGGATGATGTAGAAAAGAAGGGGT 279
             ||||||| |||||    |||||||||||| ||| || |||| ||||| ||   |||||
         228 GCTTGCGTTGTGGTAGTCCATATGATGAAAACCTGTTGGACGATGTCGA---GAAGGCCA 284

         280 CTGGCAATCAATCCACAATGGCATCTCACCTCAACGATTCTCAGGATGTCGGAATCCATG 339
             | ||| ||||||| ||||||||  | ||  | ||| | |||||||||| ||||| ||||
         285 CCGGCGATCAATCGACAATGGCTGCACATTTGAACAAGTCTCAGGATGTTGGAATTCATG 344

         340 CTAGACATATCAGTAGTGTGTCCACTGTGGATAGTGAAATGAATGATGAATATGGGAATC 399
             | |||||||||| |||||||| ||  |||||||||||||  ||| ||  ||||||||
         345 CAAGACATATCAGCAGTGTGTCTACATTGGATAGTGAAATGGCTGAAGACAATGGGAATT 404

         400 CAATTTGGAAGAATCGGGTGAAGAGCTGTAAGGATAAAGAGAACAAGAAGAAAAAGAGAA 459

Db       405 CGATTTGGAAGAACAAGGTGGAAAGTTGGAAAGAAAAGAAGAACAAGAAGAAGAAGCCTG 464

Qy       460 GTCCTAAGGCTGAAAGTGAACCAGCTCAAGTTCCTACAGAACAGCAGATGGAAGAGAAAC 519
             | |      |      |  ||| | | |  | | | | | |||||||  |||
Db       465 CAACAACTAAGGTTGAAAGAGAGGCTGAAATCTCACCTGAGCAACAAATGGAAGATAAAC 524

Qy       520 CGTCTGCAGAGGCTTCGGAGCCGCTTTGAATTCTTTATCCAATTCCACGCAACAAGCTCA 579
             || |    || | ||  |||  | |   | |  ||||| ||    | ||  ||
Db       525 CGGCACCGGATGCTTCCAGCCTCTCTCGACTATAATTCCAATCCCGAAAAGCAGACTTG 584

Qy       580 CACCATACAGAGCAGTGATCATTATGCGATTGTTCATTCTGGGCCTCTTCTTCCACTTCA 639
             ||| |||. ||  ||| |||  ||| |  | | |||||  | ||  |,||||  |
Db       585 CACCATACCGAACTTGATCATTATGAAATGATCATTCTTGGTCTTTTCTTCCATTATC 644

Qy       640 GAATAACAAATCT                                    GGCTTACTTCTGTTCATATGTGAGA 699
             ||  ||  |  |                                    ||| | ||| | || |
Db       645 GAGTAACAAA                                      TTCACTTCAGTCATATGTGAAA 704

Qy       700 TCTGGTTTGCATTT                                  CAAGTGGAATCCTGTTAATAGAG 759
             ||||:|||  ||  |||  |||   |:  :|||  | |||| ||||  || | |
Db       705 TCTGGTTTGCATTTTCTGGGTGTTTGATATATTCCCTAAGTGGTATCCTGTTAACAGGG 764

Qy       760 AAACGTATATCGAAAGGCTGTCGGCAAGGTATTAAAAGAGAGGGTGAGCCTTCTCAGCTTG 819
             |||| || || || || | || ||| | ||| ||||| |||| || | | ||||
Db       765 AAACATACATTGACAACTATCTGCAAGATATTAAAAGAGAAGGTGAACCTGATGAACTTG 824

Qy       820 CTGGTGTGGATTTTTTCGTGAGTACTGTTGATTGGCTGAAGGAACCGCCATTGATCACTG 879
             ||| | || | ||||||| | |||  |||| || |||| ||| ||||| || |||| | |
Db       825 CTGCAGTTGACTTCTTCGTGAGTACAGTTGATTGACTGAAAGAGCCTCCATTGATTACTG 884

Qy       880 CCAATACAGTCCTTTCCATCCTTGCTGTGGACTATCCCGTCGATAAAGTCTCCTGCTACG 939
             ||||||| || |||||||||||| || ||| || |||| | || | |
Db       885 CCAATACTGTGCTTTCCATGCTTGCCTTTACTACCCGGTGGATAAGGTCTCTTGTTATA 944

Qy       940 TGTCTGATGATGGTGCAGCTATGCTTTCATTTGAATCTCTTGTAGAAACAGCTGAGTTTG 999
             | |:|||  ||||:| || |||||  ||,|||||||||| ||||||||||| || ||||
Db       945 TATCTGATGATGGTGCGGCCATGCTGACATTTGAATCTCTAGTAGAAACAGCCGACTTTG 1004

Qy       1000 CAAGGAAGTGGGTTCCGTTCTGCAAAAAATTCTCAATTGAACCAAGAGCACCGGAGTTTT 1059
              ||| ||| ||| | ||||||||||| | |,|||||  | ||||  ||| ||||
Db       1005 CAAGAAAGTGGGTTCCATTCTGCAAAAAATTTTCCATTGAACCCCGGGCACCTGAGTTTT 1064

Qy       1060 ACTTCTCACAGAAAATTGATTACTTGAAAGACAAGGTTCAACCTTCTTTCGTGAAAGAAC 1119
              |||||||||.||| ||||||.|||,|||  | || || || |||| || |||||.|
Db       1065 ACTTCTCACAGAAGATTGATTACTTGAAAGATAAAGTGCAGCCCTCTTTTGTAAAAGAAC 1124

Qy       1120 GTAGAGCAATGAAAAGGGATTATGAAGAGTACAAAGTCCGAGTTAATGCCCTGGTAGCAA 1179
              |||||. ||,|||. ||||||||||||| | | | | | | | | | | | | |||
Db       1125 GTAGAGCTATGAAAAGAGATTATGAAGAGTACAAAATTCGAATCAATGCTTTAGTTGCAA 1184

Qy       1180 AGGCTCAGAAAACACTTGAAGAAGGATGGACTATGCAAGATGGAACACCTTGGCCTGGGA 1239
              |||||.||| .|||.|| |||||  |||' .|||||||||||||  |||||| .| |
Db       1185 AGGCTCAGAAAACACTTGATGAAGGATGGACAATGCAAGATGGAACTTCTTGGCCAGGAA 1244

Qy       1240 ATAACACACGTGATCACCCTGGGCATGATTCAAGTCTTCCTTGGGAAATACTGGGAGCTC 1299
              || ||  ||| ||| || | ||| ||||  | ||||||  |      ||  ,|
Db       1245 ATAACCCGCGTGATCACCCT-GGCATGATTCAAGTTTTCCTTGG--ATATAGTGGTGCTC 1301

Qy       1300 GTGACATTGAAGGAAATGAACTACCTCGTCTAGTATATGTCTCCAGGGAGAAGAGACCTG 1359
              |||||| |||||||||||||| || | | | |||| || ||||||||||||
Db       1302 GTGACATCGAAGGAAATGAACTTCCTCGACTGGTTTACGTCTCTAGAGAGAAGAGACCTG 1361

Qy       1360 GCTACCAGCACCACACAAAAAGGCTGGTGCAGAAAATGCTCTGGTGAGAGTGTCTGCAGTAC 1419
              |||||| ||||.|||||||||||||| || ||||| ||| || ||||||||||| |
Db       1362 GCTACCAACACCACAAAAAGGCTGGTGCTGAAAATGCTTTGGTTAGGGTGTCTGCAGTTC 1421

Qy       1420 TCACAAATGCTCCCTACATCCTCAATGTTGATTGTGATCACTATGTAAACAATAGCAAGG 1479
              | |||||||||||| |||||||||| |.|||| || |||||||  |||||||||||
Db       1422 TTACAAATGCTCCCTTCATCCTCAATCTTGATTGTGACCACTATGTTAACAATAGCAAGG 1481

Qy       1480 CTGTTCGAGAGGCAATGTGCATCCTGATGGACCCACAAGTAGGTCGAGATGTATGCTATG 1539
              | ||| | |||||||||||| || |||||||||||||| ||||||||||||||||||||
```

```
Db       1482 CAGTTAGGGAGGCAATGTGCTTCTTGATGGACCCACAAGTTGGTCGAGATGTATGCTATG 1541

Qy       1540 TGCAGTTCCCTCAGAGGTTTGATGGCATAGATAAGAGTGATCGCTACGCCAATCGTAACG 1599

Db       1542 TGCAGTTTCCTCAAAGATTTGATGGCATAGATAGGAGTGATCGATATGCCAATAGGAACA 1601

Qy       1600 TAGTTTTCTTTGATGTTAACATGAAAGGGTTGGATGGCATTCAAGGACCAGTATACGTAG 1659

Db       1602 CAGTTTTCTTTGATGTTAACATGAAAGGTCTTGATGGAATCCAAGGGCCAGTTTATGTGG 1661

Qy       1660 GAACTGGTTGTGTTTTCAACAGGTAAAGACTTTACGGCTAGGGCCTCCTTCTATGCCCA 1719

Db       1662 GAACAGGTTGTGTTTTCAATAGGTAAGACTTTATGGCTATGGTCCACCTTCAATGCCAA 1721

Qy       1720 GCTTACGCAAGAGAAAGGATTCTTCATCTGCTTCTCATGTTGCTGCCCCTCAAAGAAGA 1779

Db       1722 GTTTTCCCAAGTCA------TCCTCCTGATCTTGCTCGTGTTGGTGCCCGGGCAAGAAGG 1775

Qy       1780 AGCCTGCTCAAGATCCAGCTGAGGTATACAGAGATGCAAAAAGAGAGGATCTCAATGCTG 1839

Db       1776 AAC--CTAAAGATCCATCAGAGCTTTATAGGGATGCAAAACGGGAAGAACTTGATGCTG 1832

Qy       1840 CCATATTTAATCTTACAGAGATTGATAATTATGACGAGCATGAAAGGTCAATGCTGATCT 1899

Db       1833 CCATCTTTAACCTTAGGGAAATTGACAATTATGATGAGTATGAAAGATCAATGTTGATCT 1892

Qy       1900 CCCAGTTGAGCTTTGAGAAAACTTTTGGCTTATCTTCTGTCTTCATTGAGTCTACACTAA 1959

Db       1893 CTCAAACAAGCTTTGAGAAAACTTTTGGCTTATCTTCAGTCTTCATTGAATCTACACTAA 1952

Qy       1960 TGGAGAATGGAGGAGTACCCGAGTCTGCCAACTCACCACCATTCATCAAGGAAGCGATTC 2019

Db       1953 TGGAGAATGGAGGAGTGGCTGAATCTGCCAACCCTTCCACACTAATCAAGGAAGCAATTC 2012

Qy       2020 AAGTCATCGGCTGTGGCTATGAAGAGAAGACTGAATGGGGAAAACAGATTGGTTGGATAT 2079

Db       2013 ATGTCATCAGCTGTGGCTATGAAGAGAAGACTGCATGGGGGAAAAGAGATTGGATGGATAT 2072

Qy       2080 ATGGGTCAGTCACTGAGGATATCTTAAGTGGCTTCAAGATGCACTGCCGAGGATGGAGAT 2139

Db       2073 ATGGTTCAGTCACTGAGGATATCTTAAGCGGCTTCAAAATGCACTGCCGAGGATGGAGAT 2132

Qy       2140 CAATTTACTGCATGCCCGTAAGGCCTGCATTCAAAGGATCTGCACCCATCAACCTGTCTG 2199

Db       2133 CGATTTACTGCATGCCCTTAAGGCCAGCATTCAAAGGATCTGCACCCATCAATCTGTCTG 2192

Qy       2200 ATAGATTGCACCAGGTCCTCCGATGGGCTCTTGGTTCTGTGGAAATTTTCTTTAGCAGAC 2259

Db       2193 ATCGGTTGCACCAGGTTCTTCGATGGACTCTTGGATCTGTTGAAATTTTCCTAAGCAGGC 2252

Qy       2260 ACTGTCCCCTCTGGTACGGTTTGAATGTCTTAAAGGCTTGGT 2319

Db       2253 ATTGGCCTCTATGGTATGGTTTGAATGTCTTAAAGACTAGCAT 2312

Qy       2320 ATATAAACACCATTGTGTACCCATTACATCCCTCCCTCTCATTGCCTATTGCACAATTC 2379

Db       2313 ATATAAACACCATTGTCTATCCTTTCAACATCCCTTCCACTCATTGCCTATTGTTCACTAC 2372

Qy       2380 CTGCAGTTTGTCTGCTCACCGGAAAATTCATCATACCAACGCTCTCAAACCTGGCAAGCA 2439

Db       2373 CAGCAATCTGTCTTCTCACAGGAAAATTTATCATACCAACGCTCTCAAACCTGGCAAGTG 2432

Qy       2440 TGCTGTTTCTTGGCCTCTTTATCTCCATCATTGTAACTGCCGTGCTTGAGCTAAGATGGA 2499

Db       2433 TTCTCTTTCTTGGCCTTTTCCTTTCCATTATCGTGACTGCTGTTCTCGAGCTCCGATGGA 2492

Qy       2500 GCGGTGTCAGCATTGAAGATTTATGGCGTAATGAACAATTCTGGGTGATCGGAGGTGTTT 2559

Db       2493 GTGGTGTCAGCATTGAGGACTTATGGCGTAACGAGCAGTTTTGGGTCATCGGTGGCGTTT 2552

Qy       2560 CAGCCCATCTCTTTGCGGTCTTCCAGGGATTCTTAAAAATGTTGGCTGGCATCGATACGA 2619

Db       2553 CAGCCCATCTCTTTGCCGTCTTCCAAGGTTTCCTTAAGATGTTGCGGGCATTGACACCA 2612

Qy       2620 ACTTCACTGTCACAGCAAAAGCAGCCGAAGATGCAGAATTTGGGGAGCTATATATGGTCA 2679

Db       2613 ACTTTACTGTCACTGCCAAAGCAGCTGATGATGCAGATTTTGGTGAGCTCTACATTGTGA 2672

Qy       2680 AGTGGACAACACTTTTGATTCCTCCAACCACACTTCTCATTATCAATATGTCGGGT---T 2736

Db       2673 AATGGACTACACTTCTAATCCCTCCAACAACACTCCTCATCGTCAACATGGTTGGTGTCG 2732

Qy       2737 GTGCTGGATTCTCTGATGCACTCAACAAAGGATATGAAGCATGGGGGCCTCTCTTTGGCA 2796

Db       2733 TTGCCGGATTCTCCGATGCCCTCAACAAAGGGTACGAAGCTTGGGGACCACTCTTTGGCA 2792

Qy       2797 AGGTGTTCTTTGCTTTCTGGGTGATTCTTCATCTCTATCCATTCCTTAAAGGTCTAATGG 2856

Db       2793 AAGTGTTCTTTTCCTTCTGGGTCATCCTCCATCTTTATCCATTCCTCAAAGGTCTTATGG 2852

Qy       2857 GTCGCCAAAACCTAACACCAACCATTGTTGTTCTCTGGTCAGTGCTGTTGGCCTCTGTCT 2916

Db       2853 GACGCCAAAACAGGACACCAACCATTGTTGTCCTTTGGTCAGTGTTGTTGGCTTCTGTCT 2912

Qy       2917 TCTCTCTCGTTTGGGTCAAGATCAATCCATTCGTTAACAAAGTTGATAACACCTTGGTTG 2976

Db       2913 TCTCTCTTGTTTGGGTTCGGATCAACCCGTTTGTCAGCACCGCCGATAGCACCACCGTGT 2972

Qy       2977 CGGAGACCTGCATTTCCATTGATTGCTGA 3005

Db       2973 CACAGAGCTGCATTTCCATTGATTGTTGA 3001


RESULT 3
AR164434
LOCUS       AR164434                3328 bp    DNA     linear   PAT 17-OCT-2001
DEFINITION  Sequence 1 from patent US 6271443.
ACCESSION   AR164434
VERSION     AR164434.1  GI:16235654
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 3328)
  AUTHORS   Stalker,D.M., Pear,J.R. and Delmer,D.
  TITLE     Cotton and rice cellulose synthase DNA sequences
  JOURNAL   Patent: US 6271443-A 1 07-AUG-2001;
FEATURES             Location/Qualifiers
     source          1. .3328
                     /organism="unknown"
BASE COUNT      914 a    689 c    763 g    962 t
ORIGIN

  Query Match           61.6%;  Score 1989.8;  DB 6;  Length 3328;
  Best Local Similarity   80.8%;  Pred. No. 0;
  Matches 2399;  Conservative    0;  Mismatches  552;  Indels   18;  Gaps    6;

Qy         40 AAAGAGCTGGTAAAGTGGTAATAAGCAAGATGATGGAATCTGGGGCTCCTATATGCCATA 99

Db        150 AAGGAGAAAGATAAGTACTTTTTTTGAGAATGATGGAATCTGGGGTTCCTGTTTGCCACA 209

Qy        100 CCTGTGGTGAACAGGTGGGGCATGATGCAAATGGGGAGCTATTTGTGGCTTGCCATGAGT 159

Db        210 CTTGTGGTGAACATGTTGGGTTGAATGTTAATGGTGAACCTTTTGTGGCTTGCCATGAAT 269

Qy        160 GTAGCTATCCCATGTGCAAGTCTTGTTTCGAGTTTGAAATCAATGAGGGCCGGAAAGTTT 219

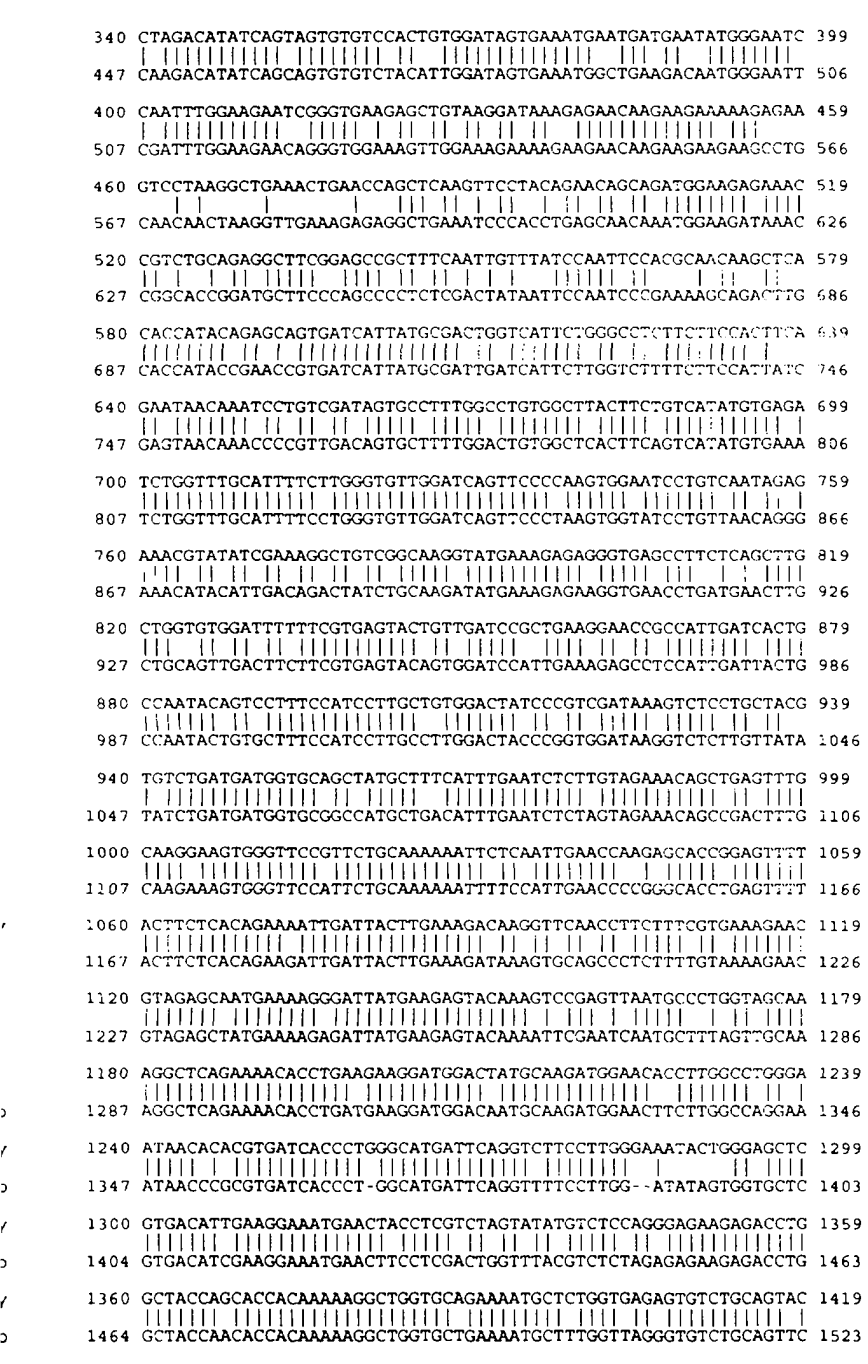
Db        270 GTAATTTCCCTATTTGTAAGAGTTGTTTTGAGTATGATCTTAAGGAAGGACGAAAAGCTT 329

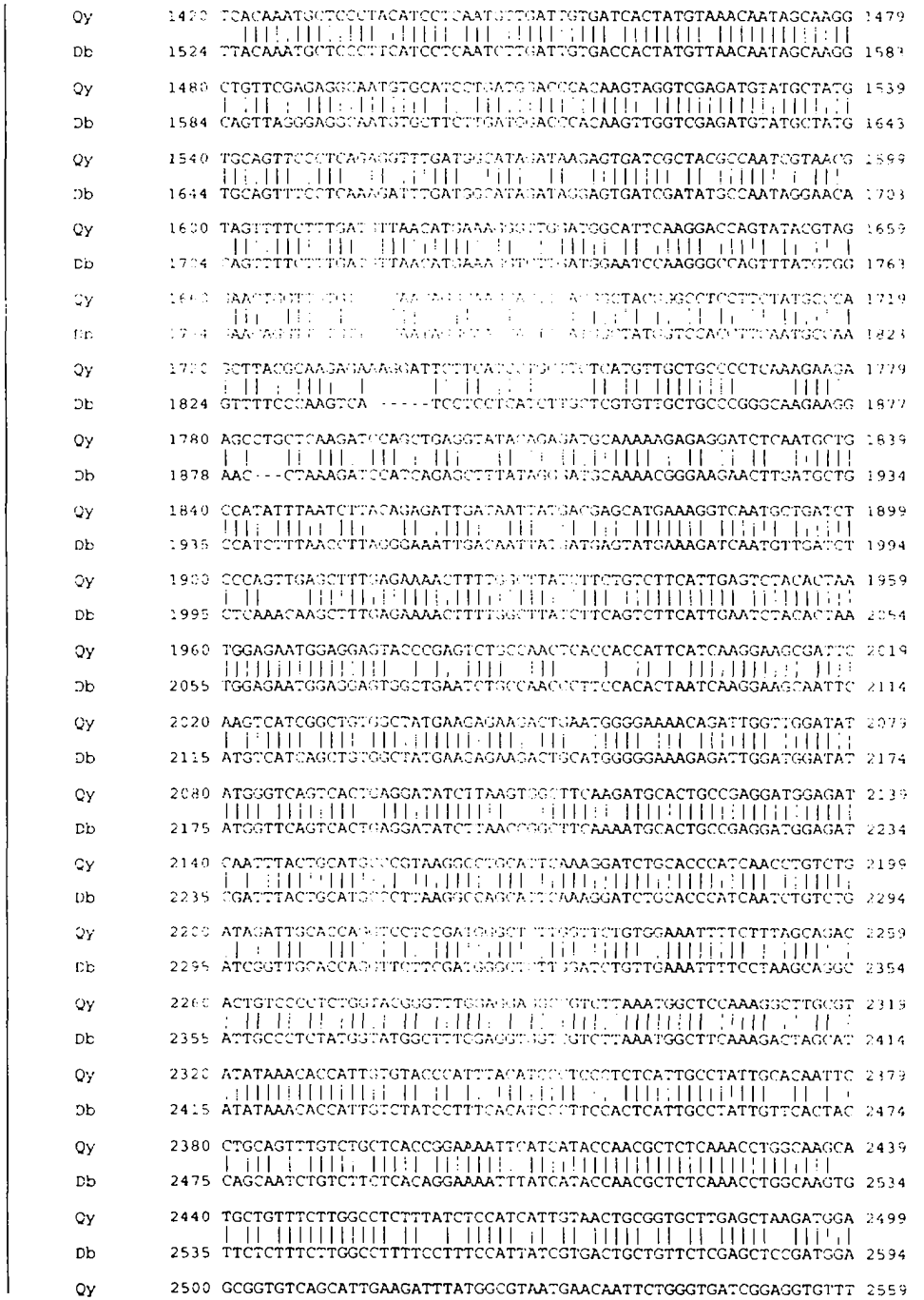
Qy        220 GCTTGCGGTGTGGCTCGCCATATGATGAGAACTTGCTGGATGATGTAGAAAAGAAGGGGT 279

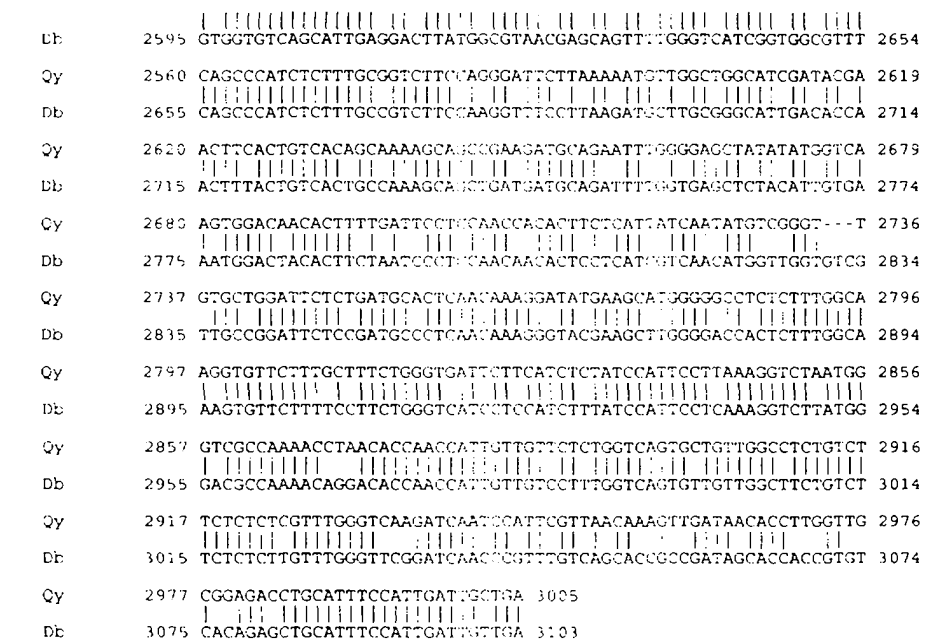
Db        330 GCTTGCGTTGTGGTAGTCCATATGATGAAAACCTGTTGGACGATGTCGA---GAAGGCCA 386

Qy        280 CTGGCAATCAATCCACAATGGCATCTCACCTCAACGATTCTCAGGATGTCGGAATCCATG 339

Db        387 CCGGCGATCAATCGACAATGGCTGCACATTTGAACAAGTCTCAGGATGTTGGAATTCATG 446
```


340 CTAGACATATCAGTAGTGTGTCCACTGTGGATAGTGAAATGAATGATGAATATGGGAATC 399
447 CAAGACATATCAGCAGTGTGTCTACATTGGATAGTGAAATGGCTGAAGACAATGGGAATT 506
400 CAATTTGGAAGAATCGGGTGAAGAGCTGTAAGGATAAAGAGAACAAGAAGAAAAAGAGAA 459
507 CGATTTGGAAGAACAGGGTGGAAAGTTGGAAAGAAAAGAAGAACAAGAAGAAGAAGCCTG 566
460 GTCCTAAGGCTGAAACTGAACCAGCTCAAGTTCCTACAGAACAGCAGATGGAAGAGAAAC 519
567 CAACAACTAAGGTTGAAAGAGAGGCTGAAATCCCACCTGAGCAACAAATGGAAGATAAAC 626
520 CGTCTGCAGAGGCTTCGGAGCCGCTTTCAATTGTTTATCCAATTCCACGCAACAAGCTCA 579
627 CGGCACCGGATGCTTCCCAGCCCCTCTCGACTATAATTCCAATCCCGAAAAGCAGACTTG 686
580 CACCATACAGAGCAGTGATCATTATGCGACTGGTCATTCTGGGCCTCTTCTTCCACTTCA 639
687 CACCATACCGAACCGTGATCATTATGCGATTGATCATTCTTGGTCTTTTCTTCCATTATC 746
640 GAATAACAAATCCTGTCGATAGTGCCTTTGGCCTGTGGCTTACTTCTGTCATATGTGAGA 699
747 GAGTAACAAACCCCGTTGACAGTGCTTTTGGACTGTGGCTCACTTCAGTCATATGTGAAA 806
700 TCTGGTTTGCATTTTCTTGGGTGTTGGATCAGTTCCCCAAGTGGAATCCTGTCAATAGAG 759
807 TCTGGTTTGCATTTTCCTGGGTGTTGGATCAGTTCCCTAAGTGGTATCCTGTTAACAGGG 866
760 AAACGTATATCGAAAGGCTGTCGGCAAGGTATGAAAGAGAGGGTGAGCCTTCTCAGCTTG 819
867 AAACATACATTGACAGACTATCTGCAAGATATGAAAGAGAAGGTGAACCTGATGAACTTG 926
820 CTGGTGTGGATTTTTTCGTGAGTACTGTTGATCCGCTGAAGGAACCGCCATTGATCACTG 879
927 CTGCAGTTGACTTCTTCGTGAGTACAGTGGATCCATTGAAAGAGCCTCCATTGATTACTG 986
880 CCAATACAGTCCTTTCCATCCTTGCTGTGGACTATCCCGTCGATAAAGTCTCCTGCTACG 939
987 CCAATACTGTGCTTTCCATCCTTGCCTTGGACTACCCGGTGGATAAGGTCTCTTGTTATA 1046
940 TGTCTGATGATGGTGCAGCTATGCTTTCATTTGAATCTCTTGTAGAAACAGCTGAGTTTG 999
1047 TATCTGATGATGGTGCGGCCATGCTGACATTTGAATCTCTAGTAGAAACAGCCGACTTTG 1106
1000 CAAGGAAGTGGGTTCCGTTCTGCAAAAAATTCTCAATTGAACCAAGAGCACCGGAGTTTT 1059
1107 CAAGAAAGTGGGTTCCATTCTGCAAAAAATTTTCCATTGAACCCCGGGCACCTGAGTTTT 1166
1060 ACTTCTCACAGAAAATTGATTACTTGAAAGACAAGGTTCAACCTTCTTTCGTGAAAGAAC 1119
1167 ACTTCTCACAGAAGATTGATTACTTGAAAGATAAAGTGCAGCCCTCTTTTGTAAAAGAAC 1226
1120 GTAGAGCAATGAAAAGGGATTATGAAGAGTACAAAGTCCGAGTTAATGCCCTGGTAGCAA 1179
1227 GTAGAGCTATGAAAAGAGATTATGAAGAGTACAAAATTCGAATCAATGCTTTAGTTGCAA 1286
1180 AGGCTCAGAAAACACCTGAAGAAGGATGGACTATGCAAGATGGAACACCTTGGCCTGGGA 1239
1287 AGGCTCAGAAAACACCTGATGAAGGATGGACAATGCAAGATGGAACTTCTTGGCCAGGAA 1346
1240 ATAACACACGTGATCACCCTGGGCATGATTCAGGTCTTCCTTGGGAAATACTGGGAGCTC 1299
1347 ATAACCCGCGTGATCACCCT-GGCATGATTCAGGTTTTCCTTGG--ATATAGTGGTGCTC 1403
1300 GTGACATTGAAGGAAATGAACTACCTCGTCTAGTATATGTCTCCAGGGAGAAGAGACCTG 1359
1404 GTGACATCGAAGGAAATGAACTTCCTCGACTGGTTTACGTCTCTAGAGAGAAGAGACCTG 1463
1360 GCTACCAGCACCACAAAAAGGCTGGTGCAGAAAATGCTCTGGTGAGAGTGTCTGCAGTAC 1419
1464 GCTACCAACACCACAAAAAGGCTGGTGCTGAAAATGCTTTGGTTAGGGTGTCTGCAGTTC 1523


Qy 1479
Db 1524
Qy 1480 CTGTTCGAGAGGCAATGTGCATCCTGATGTACCCACAAGTAGGTCGAGATGTATGCTATG 1539
Db 1584 CAGTTAGGGAGGCAATGTGCTTCTTGATGTACCCACAAGTTGGTCGAGATGTATGCTATG 1643
Qy 1540 TGCAGTTCCCTCAGAGGTTTGATGACATAGATAAGAGTGATCGCTACGCCAATCGTAACG 1599
Db 1644 TGCAGTTTCCTCAAAGATTTGATGGCATAGATAGGAGTGATCGATATGCCAATAGGAACA 1703
Qy 1600 TAGTTTTCTTTGAT TTTAACATGAAA GGTTGATGGCATTCAAGGACCAGTATACGTAG 1659
Db 1704 TAGTTTTCTTTGAT TTTAACATGAAA GTTTGATGGAATCCAAGGGCCAGTTTATGTGG 1763
Qy 1660 GAACTGTT GAA AAGTGCTACGGGCCTCCTCCTATGCCCA 1719
Db 1764 GAACAGTT AAAAGTTT ATGTCCACCTTCAATGCCAA 1823
Qy 1720 TCTTACGCAAGAGAAAAGATTCTTCATCCTGCTTCTCATGTTGCTGCCCCTCAAAGAAGA 1779
Db 1824 GTTTTCCCAAGTCA -----TCCTCCTCATCTTGCTGTGTTGCTGCCCGGGCAAGAAGG 1877
Qy 1780 AGCCTGCTCAAGATCCAGCTGAGGTATACAGAGATGCAAAAAGAGAGGATCTCAATGCTG 1839
Db 1878 AAC---CTAAAGATCCATCAGAGCTTTATAGCGATGCAAAACGGGAAGAACTTGATGCTG 1934
Qy 1840 CCATATTTAATCTTACAGAGATTGATAATTATGACGAGCATGAAAGGTCAATGCTGATCT 1899
Db 1935 CCATCTTTAAGCTTAGGAAATTGACAATTAT GATGAGTATGAAAGATCAATGTTGATCT 1994
Qy 1900 CCCAGTTGAGCTTTGAGAAAACTTTTGGCTTATCTTCTGTCTTCATTGAGTCTACACTAA 1959
Db 1995 CTCAAACAAGCTTTGAGAAAACTTTTGGCTTATCTTCAGTCTTCATTGAATCTACACTAA 2054
Qy 1960 TGGAGAATGGAGGAGTACCCGAGTCTGCCAACTCACCACCATTCATCAAGGAAGCGATTC 2019
Db 2055 TGGAGAATGGAGGAGTGGCTGAATCTGCCAACCCTTCCACACTAATCAAGGAAGCAATTC 2114
Qy 2020 AAGTCATCGGCTGTGGCTATGAACAGAAGACTGAATGGGGAAAACAGATTGGTTGGATAT 2079
Db 2115 ATGTCATCAGCTGTGGCTATGAACAGAAGACTGCATGGGGGAAAGAGATTGGATGGATAT 2174
Qy 2080 ATGGGTCAGTCACTCAGGATATCTTAAGTGGCTTCAAGATGCACTGCCGAGGATGGAGAT 2139
Db 2175 ATGGTTCAGTCACTGAGGATATCTTAACCGGCTTCAAAATGCACTGCCGAGGATGGAGAT 2234
Qy 2140 CAATTTACTGCATGCCCGTAAGGCCTGCATTCAAAGGATCTGCACCCATCAACCTGTCTG 2199
Db 2235 CGATTTACTGCATGCCCTTAAGGCCAGCATTCAAAGGATCTGCACCCATCAATCTGTCTG 2294
Qy 2200 ATAGATTGCAGCA TCCTCCGATGGGCT TTGGTTCTGTGGAAATTTTCTTTAGCAGAC 2259
Db 2295 ATCGGTTGCAACCA TTCTTCGATGGGCT TTGGATCTGTTGAAATTTTCCTAAGCAGGC 2354
Qy 2260 ACTGTCCCCTCTGGTACGGGTTTGGA GCTGTCTTAAATGGCTCCAAAGGCTTGCGT 2319
Db 2355 ATTGCCCTCTATGGTATGGCTTTCGAGGT GCTGTCTTAAATGGCTTCAAAGACTAGCAT 2414
Qy 2320 ATATAAACACCATTGTGTACCCATTTACATCCCTCCCTCTCATTGCCTATTGCACAATTC 2379
Db 2415 ATATAAACACCATTGTCTATCCTTTCACATCCCTTCCACTCATTGCCTATTGTTCACTAC 2474
Qy 2380 CTGCAGTTTGTCTGCTCACCGGAAAATTCATCATACCAACGCTCTCAAACCTGGCAAGCA 2439
Db 2475 CAGCAATCTGTCTTCTCACAGGAAAATTTATCATACCAACGCTCTCAAACCTGGCAAGTG 2534
Qy 2440 TGCTGTTTCTTGGCCTCTTTATCTCCATCATTGTAACTGCGGTGCTTGAGCTAAGATGGA 2499
Db 2535 TTCTCTTTCTTGGCCTTTTCCTTTCCATTATCGTGACTGCTGTTCTCGAGCTCCGATGGA 2594
Qy 2500 GCGGTGTCAGCATTGAAGATTTATGGCGTAATGAACAATTCTGGGTGATCGGAGGTGTTT 2559



RESULT 4
AY099636
LOCUS AY099636 3109 bp mRNA linear PLN 06-MAY-2002
DEFINITION Arabidopsis thaliana cellulose synthase like protein (At4g18780) mRNA, complete cds.
ACCESSION AY099636
VERSION AY099636.1 GI:20466339
KEYWORDS FLI_CDNA.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis
REFERENCE 1 (bases 1 to 3109)
AUTHORS Nguyen,M., Karlin-Neumann,G., Lam,B., Miranda,M., Palm,C.J., Bowser,L., Jones,T., Bank,C., Chen,H., Cheuk,R., Chung,M.K., Hayashizaki,Y., Ishida,J., Kamiya,A., Kawai,J., Kim,C., Lin,J., Liu,S.X., Narusaka,M., Pham,P.K., Sakano,H., Sakurai,T., Satou,M., Seki,M., Shinn,P., Yamada,K., Shinozaki,K., Ecker,J., Theologis,A. and Davis,R.W.
TITLE Direct Submission
JOURNAL Submitted (24-APR-2002) DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA
COMMENT e-mail for correspondence: arabidsequence@genome.stanford.edu

RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RAFL cDNAs (RAFL cDNA : "RIKEN Arabidopsis Full-Length cDNA"). Seki,M., Narusaka,M., Ishida,J., Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J., Hayashizaki,Y. and Shinozaki,K.

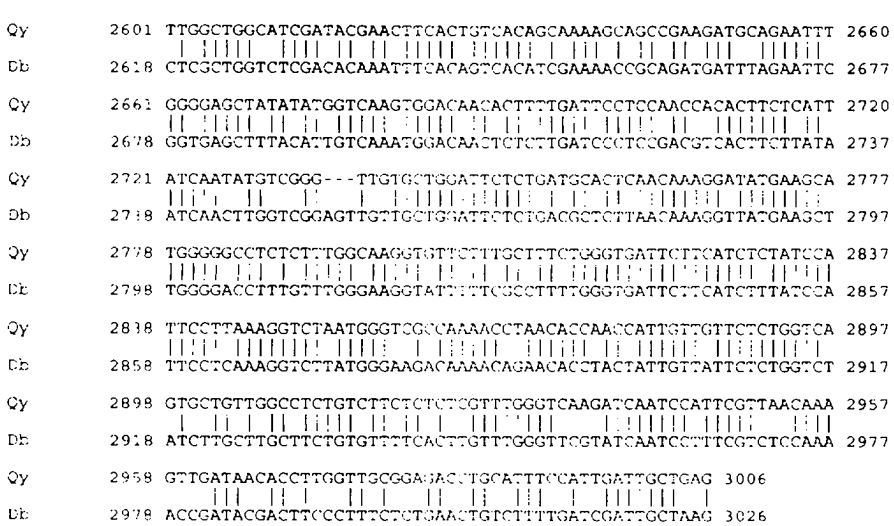
The Salk, Stanford, PGEC (SSP) Consortium members carried out the sequencing and annotation of the RAFL cDNAs: Nguyen,M., Southwick,A., Karlin-Neumann,G., Lam,B., Miranda,M., Palm,C.J., Bowser,L., Jones,T., Bank,C., Chen,H., Cheuk,R., Chung,M.K., Kim,C., Lin,J., Liu,S.X., Pham,P.K., Sakano,H., Shinn,P., Yamada,K., Ecker,J., Theologis,A. and Davis,R.W.

Nguyen,M, (SSP/Stanford) and Seki,M. (RIKEN GSC) contributed equally to this work. Shinozaki,K. (RIKEN GSC) and Davis,R.W. (SSP/Stanford) contributed equally to this work as PIs.

FEATURES Location/Qualifiers
source 1..3109
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/chromosome="4"
/clone="RAFL09-65-M12 (R24904)"
/note="This clone is in pBluescript
ecotype: Columbia"
gene 1..3109
/gene="At4g18780"
/note="synonym: F28A21.190"
CDS 68..3025
/gene="At4g18780"
/codon_start=1
/product="cellulose synthase-like protein"
/protein_id="AAM20487.1"
/db_xref="GI:20466340"
/translation="MMESRSPICNTCGEEIGVKSNGEFFVACHECSFPICKACLEYEF
KEGRRICLRCGNPYDENVFDDVETKTSKTQSIVPTQTNNTSQDSGIHARHISTVSTID
SELNDEYGNPIWKNRVESWKDKKDKKSKKKKKDPKATKAEQHEAQIPTQQHMEDTPPN
TESGATDVLSVVIPIPRTKITSYRIVIIMRLIILALFFNYRITHPVDSAYGLWLTSVI
CEIWFAVSWVLDQFPKWSPINRETYIDRLSARFEREGEQSQLAAVDFFVSTVDPLKEP
PLITANTVLSILALDYPVDKVSCYVSDDGAAMLSFESLVETADFARKWVPFCKKYSIE
PRAPEFYFSLKIDYLRDKVQPSFVKERRAMKRDYEEFKIRMNALVAKAQKTPEEGWTM
QDGTSWPGNNTRDHPGMIQVFLGYSGARDIEGNELPRLVYVSREKRPGYQHHKKAGAE
NALVRVSAVLTNAPFILNLDCDHYVNNSKAVREAMCFLMDPVVGQDVCFVQFPQRFDG
IDKSDRYANRNIVFFDVNMRGLDGIQGPVYVGTGTVFRRQALYGYSPPSKPRILPQSS
SSSCCCLTKKKQPQDPSEIYKDAKREELDAAIFNLGDLDNYDEYDRSMLISQTSFEKT
FGLSTVFIESTLMENGGVPDSVNPSTLIKEAIHVISCGYEEKTEWGKEIGWIYGSITE
DILTGFKMHCRGWRSIYCMPLRPAFKGSAPINLSDRLHQVLRWALGSVEIFLSRHCPL
WYGCSGGRLKLLQRLAYINTIVYPFTSLPLVAYCTLPAICLLTGKFIIPTLSNLASML
FLGLFISIILTSVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQGFLKMLAGLDT
NFTVTSKTADDEFGELYIVKWTTLLIPPTSLLIINLVGVVAGFSDALNKGYEAWGPL
FGKVFFAFWVILHLYPFLKGLMGRQNRTPTIVILWSILLASVFSLVWVRINPFVSKTD
TTSLSLNCLLIDC"
BASE COUNT 859 a 631 c 687 g 932 t
ORIGIN

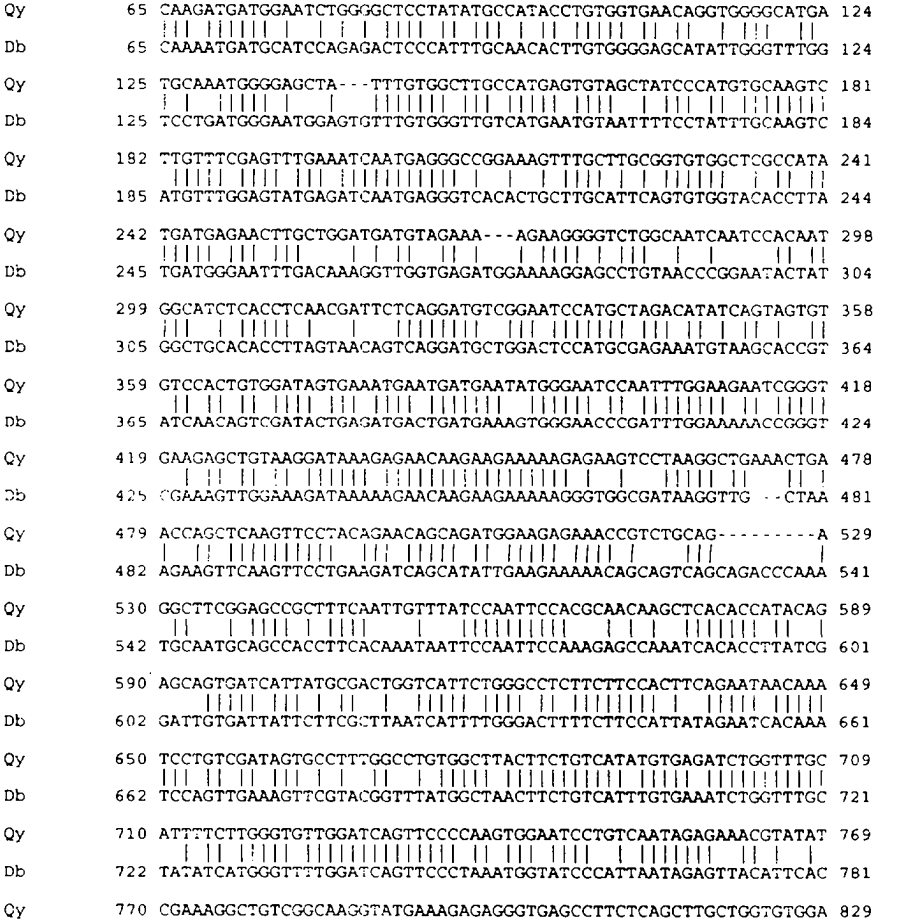
Query Match 51.9%; Score 1677; DB 8; Length 3109;
Best Local Similarity 74.4%; Pred. No. 0;
Matches 2224; Conservative 0; Mismatches 720; Indels 45; Gaps 7;





RESULT 5
AF323039
LOCUS AF323039 3145 bp mRNA linear PLN 25-FEB-2002
DEFINITION Zinnia elegans cellulose synthase CesA-1 mRNA, complete cds.
ACCESSION AF323039
VERSION AF323039.2 GI:18875453
KEYWORDS .
SOURCE Zinnia elegans
ORGANISM Zinnia elegans
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; campanulids; Asterales; Asteraceae; Asteroideae; Heliantheae; Zinnia.
REFERENCE 1 (bases 1 to 3145)
AUTHORS Haigler,C.H., Ivanova-Datcheva,M., Hogan,P.S., Salnikov,V.V., Hwang,S., Martin,K. and Delmer,D.P.
TITLE Carbon partitioning to cellulose synthesis
JOURNAL Plant Mol. Biol. 47 (1-2), 29-51 (2001)
MEDLINE 21438255
PUBMED 11554477
REFERENCE 2 (bases 1 to 3145)
AUTHORS Hwang,S. and Haigler,C.H.
TITLE Direct Submission
JOURNAL Submitted (22-NOV-2000) Biological Sciences, Texas Tech University, Box 43131, Lubbock, TX 79409, USA
REFERENCE 3 (bases 1 to 3145)
AUTHORS Hwang,S. and Haigler,C.H.
TITLE Direct Submission
JOURNAL Submitted (25-FEB-2002) Biological Sciences, Texas Tech University, Box 43131, Lubbock, TX 79409, USA
REMARK Sequence update by submitter
COMMENT On Feb 25, 2002 this sequence version replaced gi:13021930.
FEATURES Location/Qualifiers
source 1..3145
/organism="Zinnia elegans"
/mol_type="mRNA"
/db_xref="taxon:34245"
CDS 69..3005
/codon_start=1
/product="cellulose synthase CesA 1"
/protein_id="AAK11588.2"
/db_xref="GI:18875454"
/translation="MMHPETPICNTCGEHIGFGPDGNGVFVGCHECNFPICKSCLEYE
INEGHTACIQCGTPYDGNLTKVGEMEKEPVTRNTMAAHLSNSQDAGLHARNVSTVSTV
DTEMTDESGNPIWKNRVESWKDKKNKKKKGGDKVAKEVQVPEDQHIEEKQQSADPNAM
QPPSQIIPIPKSQITPYRIVIILRLIILGLFFHYRITNPVESSYGLWLTSVICEIWFA
ISWVLDQFPKWYPINRVTFTDELSARYEREGEPSELAAVDFFVSTVDPLKEPPLITAN
TVLSILAVDYPVEKVSCYVSDDGAAMLTFESLVETAEFARKWVPFCKKFSIEPRAPEF
YFSQKIDYLKDKVQPSFVKERRAMKRDYEEYKVRVNALVAKAQKTPEEGWTMQDGTPW
PGNNPRDHPGMIQVFLGHSGAHDIEGNELPRLVYVSREKRPGYQHHKKAGAENALVRV
SAVLTNAPYILNLDCDHYVNNSKAVREAMCFMMDPQVGRDVCYIQFPQRFDGIDRSDR
YANRNTVFFDVNMKGLDGIQGPVYVGTGCVFYRQALYGYGPQSLPTLPSPSSSSSCCC
CGPKKPKKDLEEFKRDARRDDLNAAIFNLKEIESYDDYERSLLISQMSFEKTFGMSSV
FIESTLMENGGLAESANPATMINEAIHVISCGYEEKTAWGKEIGWIYGSVTEDILTGF
KMHCRGWRSIYCMPVRPAFKGSAPINLSDRLHQVLRWALGSVEIFLSRHCPLWYGWGG
GRLKLLQRLAYINTIVYPFTSLPLVAYCTLPAICLLTGKFIIPTLSNIAAVWFLGLFL
SIITTSVLEIRWSGVSIEELWRNEQFWVIGGVSAHLFAVFQGFLKMLAGVDTNFTVTA
KAADDQEFGELYMIKWTTVLIPPTTLLVLNLVGVVAGFSDALNKGYEAWGPLFGKVFF
AFWVILHLYPFLKGLMGRQNRTPTIVILWSVLLASVFSLVWVKIDPFVSKGDSNLTQG
CIAIDC"
BASE COUNT 873 a 582 c 746 g 944 t
ORIGIN

Query Match 51.6%; Score 1666.8; DB 8; Length 3145;
Best Local Similarity 74.9%; Pred. No. 0;
Matches 2217; Conservative 0; Mismatches 707; Indels 36; Gaps 9;





```
Qy       2987 CATTTCCATTGATTGCTGAG 3006
              |||| | ||||||| ||||
Db       2987 CATTGCAATTGATTGTTGAG 3006


RESULT 6
AF267742
LOCUS       AF267742               2958 bp    mRNA    linear   PLN 16-FEB-2001
DEFINITION  Arabidopsis thaliana cellulose synthase catalytic subunit (IRX1)
            mRNA, complete cds.
ACCESSION   AF267742
VERSION     AF267742.1  GI:12836996
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 2958)
  AUTHORS   Taylor,N.G., Laurie,S. and Turner,S.R.
  TITLE     Multiple cellulose synthase catalytic subunits are required for
            cellulose synthesis in Arabidopsis
  JOURNAL   Plant Cell 12 (12), 2529 2540 (2000)
  MEDLINE   21066939
   PUBMED   11148295
REFERENCE   2  (bases 1 to 2958)
  AUTHORS   Taylor,N.G. and Turner,S.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-MAY-2000) Biological Sciences, University of
            Manchester, 3.614 Stopford Building, Oxford Road, Manchester M13
            9PT, UK
FEATURES             Location/Qualifiers
     source          1. .2958
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
     gene            1. .2958
                     /gene="IRX1"
     CDS             1. .2958
                     /gene="IRX1"
                     /codon_start=1
                     /product="cellulose synthase catalytic subunit"
                     /protein_id="AAK08700.1"
                     /db_xref="GI:12836997"
                     /translation="MMESRSPICNTCGEEIGAKSNGEFFVACHECSFPICKACLEYEF
                     KEGRRICLRCGNPYDENVFDDVETKTSKTQSIVPTQPNNTSQDSGIHAPHISTVSTID
                     SELNDEYGNPIWKNRVESWKDKKDKKSKKKKKDEPKATKAEQDAQVPTQQHMEDMPPN
                     TESGATDVLSVVIPIIPTEITSYRIVIIMRLIILALFFNYRITHPVDSAYGLWLTSVI
                     CEIWFAVSWVLDQFPKWSPINRETYIDRLSARYEREGEQSQLAAVDFFVSTVDPLKEP
                     PLITANTVLSILAVDYPV.KVSTYVSDDGAAMLTFEALSETAEFARKWVPFCKKYSIE
                     PRAPEFYFSLKIDYLKDKVQPSFVKERRAMKREYEEFKIRMNALVAKAQKTPEEGWTM
                     QDGTSWPGNNTRDHPGMIQVFLGYSGARDTEGNELPRLVYVSREKRPGYQHHKKAGAE
                     NALVRVSAVLTNAPFILNLDCDHYVNNSKAVREAMCFLMDPVVGQDVCFVQFPQRFDG
                     IDKSDRYANRNIVFFDVNMRGLDGIQGPVYVGTGTVFRRQALYGYSPPSKPRILPQSS
                     SSSCCCLTKKKQPQDPAEIYKDAKREELDAAIFNLGDLDNYDEYDRSMLISQTSFEKT
                     FGLSTVFIESTLMENGGVPDSVNPSTLIKEAIHVISCGYEEKTEWGKEIGWIYGSITE
                     DILTGFKMHCRGWRSIYCMPLRPAFKGSAPINLSDRLHQVLRWALGSVEIFLSRHCPL
                     WYGCSGGRLKLLQRLAYINTIVYPFTSLPLVAYCTLPAICLLTGKFIIPTLSNLASML
                     FLGLFISIILTSVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQGFLKMLAGLDT
                     NFTVTSKTADDLEFGELYIVKWTTLLIPPTSLLIINLVGVVAGFSDALNKGYEAWGPL
                     FGKVFFAFWVILHLYPFLKGLMGRQNRTPTIVILWSILLASVFSLVWVRINPFVSKTD
                     TTSLSLNCLLIDC"
BASE COUNT      818 a    610 c    658 g    872 t
ORIGIN

  Query Match           51.5%;  Score 1666;  DB 8;  Length 2958;
  Best Local Similarity   74.4%;  Pred. No. 0;
  Matches 2210;  Conservative    0;  Mismatches  715;  Indels   45;  Gaps    7;

Qy         69 ATGATGGAATCTGGGGCTCCTATATGTCATACCTGTGGTGAACAGGTGGGGCATGATGCA 128
              |||||||| ||| || |||| | ||| | || ||||||||| || | ||     | ||
Db          1 ATGATGGAGTCTAGGTCTCCCATCTGCAACACTTGTGGTGAAGAGATTGGTGCAAAATCA 60

Qy        129 AATGGGGAGCTATTTGTGGCTTGCCATGAGTGTAGCTATCCCATGTGCAAGTCTTGTTTC 188
              || || ||| | |||||| |||| |||||||||||| |  || || |||||  ||||| |
Db         61 AACGGAGAGTTCTTTGTGGCTTGTCATGAGTGTAGTTTCCCGATCTGCAAAGCTTGTCTT 120

Qy        189 GAGTTTGAAATCAATGAGGGCCGGAAAGTTTGCTTGCGGTGTGGCTCGCCATATGATGAG 248
              |||| |||| |||| || | || | | |||||||||| || |||   || || ||||||
Db        121 GAGTATGAATTCAAAGAAGGTCGAAGAATTTGCTTGCGTTGCGGCAATCCTTACGATGAG 180

Qy        249 AACTTGCTGGATGATGTAGAAAAGAAGGGGTCTGGCAATCAATCCACAATGGCATCTCA- 307
              ||  || | |||||||| | | ||| ||| | ||||||||  |  || | ||
Db        181 AATGTGTTTGATGATGTTGAGACAAAGACATCTAAAACTCAATCCATTGTTCCAACACAG 240

Qy        308 --CCTCAACGATTCTCAGGATGTCGGAATCCATGCTAGACATATCAGTAGTGTGTCCACT 365
                |   ||| |||||| ||  || || ||||||||||||||| |||| || || ||
Db        241 ACCAATAACACTTCTCAGGATTCAGGGATTCATGCTAGACATATAAGTACAGTCTCAACA 300

Qy        366 GTGGATAGTGAAATGAATGATGAATATGGGAATCCAATTTGGAAGAATCGGGTGAAGAGC 425
              | || |||||| |||||||||||||||| |||||||||||||||||  | ||| |||||
Db        301 ATAGACAGTGAACTGAATGATGAATATGGCAATCCAATTTGGAAGAACAGAGTGGAGAGC 360

Qy        426 TGTAAGGATAAAGAGAACAAGAAGAAAAAG----------------AGAAGTCCTAAG 467
              || || || ||  | |||||||| |||                   | |   | ||
Db        361 TGGAAAGACAAGAAAGACAAGAAGAGCAAGAAGAAGAAGAAAGATCCAAAAGCAACAAAA 420

Qy        468 GCTGAAACTGAACCAGCTCAAGTTCCTACAGAACAGCAGATGGAAGA---------GAAA 518
              ||||||    ||  ||||| |||||||  |||| || ||||||||          |||
Db        421 GCTGAACAACAAGACGCTCAGGTTCCTACCCAACAACACATGGAAGATATGCCACCGAAC 480

Qy        519 CCGTCTGCAGAGGCTTCGGAGCCGCTTTCAATTGTTTATCCAATTCCACGCAACAAGCTC 578
              |    | |  ||| | ||   ||||||  |||  ||| || ||  | |  ||| ||
Db        481 ACAGAATCTGGTGCTACAGATGTGCTTTCGGTTGTGATTCCTATCCCGAGAACAAAGATC 540

Qy        579 ACACCATACAGAGCAGTGATCATTATGCGACTGGTCATTCTGGGCCTCTTCTTCCACTTC 638
              ||  |||||||    || ||||| |||||  || ||||  |||  || || ||  ||| |
Db        541 ACTTCATACAGGATCGTTATCATCATGCGGTTGATCATCTTGGCTCTGTTTTTTAACTAC 600

Qy        639 AGAATAACAAATCCTGTCGATAGTGCCTTTGGCCTGTGGCTTACTTCTGTCATATGTGAG 698
              | || ||  |||||||||||||||| |  ||  | ||||| || ||||| |||||||||
Db        601 CGTATCACTCATCCTGTCGATAGTGCTTACGGTTTATGGCTAACATCTGTGATATGTGAG 660

Qy        699 ATCTGGTTTGCATTTTCTTGGGTGTTGGATCAGTTCCCCAAGTGGAATCCTGTCAATAGA 758
              || |||||||| |||||||||||||||||||| || ||||||  |||| | || ||
Db        661 ATATGGTTTGCTGTTTCTTGGGTGTTGGATCAGTTTCCTAAGTGGTCTCCTATTAACCGA 720

Qy        759 GAAACGTATATCGAAAGGCTGTCGGCAAGGTATGAAAGAGAGGGTGAGCCTTCTCAGCTT 818
              ||||| || |||||  || | || |||||  | |||||||| || ||||  || ||||||
Db        721 GAAACTTACATCGATCGGTTATCCGCAAGATTTGAAAGAGAAGGAGAGCAATCACAGCTT 780

Qy        819 GCTGGTGTGGATTTTTTCGTGAGTACTGTTGATCCGCTGAAGGAACCGCCATTGATCACT 878
              ||   ||| ||||| || || ||||| || |||||| ||||||| || || |||||||||
Db        781 GCAGCTGTAGATTTCTTTGTTAGTACGGTAGATCCGTTGAAGGAGCCACCTTTGATCACT 840

Qy        879 GCCAATACAGTCCTTTCCATCCTTGCTGTGGACTATCCCGTCGATAAAGTCTCCTGCTAC 938
              || || || || ||||| |||||  ||  | || ||||| || ||||||||||| |||||
Db        841 GCAAACACGGTTCTTTCGATCCTCGCGCTTGATTATCCGGTGGATAAAGTCTCTTGCTAT 900

Qy        939 GTGTCTGATGATGGTGCAGCTATGCTTTCATTTGAATCTCTTGTAGAAACAGCTGAGTTT 998
              || |||||||||||||| || |||||||| ||||| ||| | || || ||||| || |||
Db        901 GTATCTGATGATGGTGCTGCAATGCTTTCTTTTGAGTCTTTGGTTGAGACAGCAGATTTT 960

Qy        999 GCAAGGAAGTGGGTTCCGTTCTGCAAAAAATTCTCAATTGAACCAAGAGCACCGGAGTTT 1058
              || ||||| ||||| || ||||| ||||| | ||| || || ||| |||| || ||||||
Db        961 GCTAGGAAATGGGTACCTTTCTGTAAAAAGTACTCCATCGAGCCACGAGCTCCCGAGTTT 1020

Qy       1059 TACTTCTCACAGAAAATTGATTACTTGAAAGACAAGGTTCAACCTTCTTTCGTGAAAGAA 1118
              |||||||| |  ||||| |||||||||  || || |||||||||||||| |||||||||
Db       1021 TACTTCTCGCTTAAAATCGATTACTTGAGGGATAAAGTTCAACCTTCTTTTGTGAAAGAA 1080

Qy       1119 CGTAGAGCAATGAAAAGGGATTATGAAGAGTACAAAGTCCGAGTTAATGCCCTGGTAGCA 1178
              |||||||| |||||||| ||||||||||||||  ||| |  || | |||||  | || ||
Db       1081 CGTAGAGCCATGAAAAGAGATTATGAAGAGTTTAAAATAAGAATGAATGCTTTAGTCGCC 1140
```

```
1179 AAGGCTCAGAAAACACCTGAAGAAGGATGGACTATGCAAGATGGAACACCTTGGCCTGGG 1238
     |||||||| || ||||| |||||||||||||| |||||||||||||||| ||||||| ||
1141 AAGGCTCAAAAGACACCAGAAGAAGGATGGACAATGCAAGATGGAACATCTTGGCCCGGG 1200

1239 AATAACACACGTGATCACCCTGGGCATGATTCAGGTCTTCCTTGGGAAATACTGGGAGCT 1298
     || ||||| ||||| || || ||, |||||||||||| || ||||| | | || |||
1201 AACAACACTCGTGACCATCCCGGG-ATGATTCAGGTTTTTCTTGGATATAGC--GGTGCT 1257

1299 CGTGACATTGAAGGAAATGAACTACCTCGTCTAGTATATGTCTCCAGGGAGAAGAGACCT 1358
     || |||||||||||||||||||| || | |||| || ||||| |||||||||||||||
1258 CGCGACATTGAAGGAAATGAACTTCCAAGATTAGTTTACGTCTCTAGGGAGAAGAGACCT 1317

1359 GGCTACCAGCACCACAAAAAGGCTGGTGCAGAAAATGCTCTGGTGAGAGTGTCTGCAGTA 1418
     || || ||||| |||||||||||| || ||||| || || |||||| |||||||| ||
1318 GGTTATCAGCATCACAAAAAGGCCGGGGCAGAGAACGCATTGGTGAGGGTGTCTGCGGTT 1377

1419 CTCACAAATGCTCCCTACATCCTCAATGTTGATTGTGATCACTATGTAAACAATAGCAAG 1478
     | || |||||||| | ||| || || ||||||||||||||| || ||||||||||
1378 TTAACGAATGCTCCATTCATTCTTAACCTTGATTGTGATCACTACGTCAACAATAGCAAA 1437

1479 GCTGTTCGAGAGGCAATGTGCATCCTGATGGACCCACAAGTAGGTCGAGATGTATGCTAT 1538
     || || || || |||||||| | | ||||| || |||| ||| | |||| |
1438 GCCGTGCGTGAAGCAATGTGCTTTTTAATGGATCCTGTTGTTGGTCAAGACGTTTGCTTT 1497

1539 GTGCAGTTCCCTCAGAGGTTTGATGGCATAGATAAGAGTGATCGCTACGCCAATCGTAAC 1598
     || |||||||| ||||| |||||||| || || |||||||||| || || || ||| |||
1498 GTTCAGTTCCCACAGAGATTTGATGGAATCGACAAGAGTGATCGATATGCTAACCGCAAC 1557

1599 GTAGTTTTCTTTGATGTTAACATGAAAGGGTTGGATGGCATTCAAGGACCAGTATACGTA 1658
     | |||||||| |||||||| |||| |||| | ||||| |||||||| |||| || ||
1558 ATTGTTTTCTTCGATGTTAATATGAGAGGGCTTGATGGGATTCAAGGTCCAGTTTATGTT 1617

1659 GGAACTGGTTGTGTTTTCAACAGGCAAGCACTTTACGGCTACGGGCCTCCTTCTATGCCC 1718
     || || ||| ||| || |  || ||||||||||||||| ||| | || ||||| |  ||
1618 GGTACAGGTACTGTCTTTAGAAGACAAGCACTTTACGGATACAGTCCACCTTCAAAACCG 1677

1719 AGCTTACGCAAGAGAAAGGATTCTTCATCCTGCTTCTCATGTTGCTGCCCCTCAAAGAAG 1778
     || |       | |          || || |  | || |||||||| |   | ||||||
1678 AGGATTTTACCGCAA---------TCTTCATCATCGTCGTGTTGCTGTCTAACCAAGAAG 1728

1779 AAGCCTGCTCAAGATCCAGCTGAGGTATACAGAGATGCAAAAAGAGAGGATCTCAATGCT 1838
     || |   | |||||||| || ||| | |||| ||| ||||| |||| || ||  ||||
1729 AAACAACCCCAAGATCCTGCCGAGATTTACAAAGACGCAAAGCGAGAAGAACTTGATGCA 1788

1839 GCCATATTTAATCTTACAGAGATTGATAATTATGACGAGCATGAAAGGTCAATGCTGATC 1898
     || || |||||||||   ||  | || || || || ||  | || || ||||||||||
1789 GCAATCTTTAATCTTGGGGACCTAGACAACTACGATGAATACGACAGATCAATGCTGATT 1848

1899 TCCCAGTTGAGCTTTGAGAAAACTTTTGGCTTATCTTCTGTCTTCATTGAGTCTACACTA 1958
     || ||    |||||||||||||||| |||||  | ||| | || ||||| ||||||||  ||
1849 TCACAAACAAGCTTTGAGAAAACGTTTGGTCTCTCTACGGTTTTCATCGAGTCTACTCTT 1908

1959 ATGGAGAATGGAGGAGTACCCGAGTCTGCCAACTCACCACCATTCATCAAGGAAGCGATT 2018
     ||||| ||||| || || ||||| |||| ||| |  || |  | ||||| ||||| |||
1909 ATGGAAAATGGCGGTGTTCCCGACTCTGTAAACCCGTCAACGCTTATCAAAGAAGCTATT 1968

2019 CAAGTCATCGGCTGTGGCTATGAAGAGAAGACTGAATGGGGAAAACAGATTGGTTGGATA 2078
     || ||||||  | ||||| || |||||||| || |||||||||||| | || || |||||
1969 CATGTCATTAGTTGTGGATACGAAGAGAAAACCGAATGGGGAAAAGAAATCGGATGGATT 2028

2079 TATGGGTCAGTCACTGAGGATATCTTAAGTGGCTTCAAGATGCACTGCCGAGGATGGAGA 2138
     || ||||  |||| || || || || | || |||||||||| || || ||||||||||
2029 TACGGGTCGATCACGGAAGACATTTTGACGGGTTTCAAGATGCATTGTCGCGGATGGAGG 2088

2139 TCAATTTACTGCATGCCCGTAAGGCCTGCATTCAAAGGATCTGCACCCATCAACCTGTCT 2198
     || |||||||| |||||  |||| || || || ||||| |||||||| ||||| || ||
2089 TCGATTTACTGTATGCCATTAAGACCAGCTTTTAAAGGTTCTGCACCAATCAATCTATCA 2148

2199 GATAGATTGCACCAGGTCCTCCGATGGGCTCTTGGTTCTGTGGAAATTTTCTTTAGCAGA 2258
     |||||  | |||||||| || || |||||||||||| || || || || ||| ||||| ||
2149 GATAGGCTTCACCAGGTTCTACGTTGGGCTCTTGGCTCGGTTGAGATCTTCCTTAGCCGA 2208

2259 CACTGTCCCCTCTGGTACGGGTTTGGAGGAGGCCGTCTTAAATGGCTCCAAAGGCTTGCG 2318
```

```
Db        2209 CATTGTCCTTTGTGGTACGGTTGCAGCGGAGGCGCCTCAAGTTACTCCAGAGATTAGCT 2268

Qy        2319 TATATAAACACCATTGTGTACCGATTACATCCCTCCCTCTCATTGCCTATTGCACAATT 2378
               ||||||||||| |||| ||| ||| | || | |||| |||| || || || ||
Db        2269 TATATAAACACTATTGTTACCGATTCACTTCTTTGCCTCTTGTTGCTTACTGTACTGTT 2328

Qy        2379 CCTGCAGTTTGTCTGCTCACCGGAAAATTCATCATACCAACGCTCTCAAACCTGGCAAGC 2438
               || ||  ||| |  || ||| |||| || | |||||| |||||||  ||||
Db        2329 CCAGCTATTTGCTTTTTACCGGCAAATTTATCATCCCAACGCTATCAAACCTAGCAAGC 2388

Qy        2439 ATGCTGTTTCTTGG TCTTTATCTCCATCATTGTAACTGCGGTGCTTGAGCTAAGATGG 2498
               |||||||||| | |  ||| |||| ||  ||  |||     || || |||||  ||||
Db        2389 ATGCTGTTTCTAGGTCTTTTATATCAATCATCTTAACGAGTGTCCTCGAGCTTCGATGG 2448

Qy        2499 AGCGGTGTCAGCATCGAAGATTTATGGTGTAATGAACAATTCTGGGTGATCGGAGGTGTT 2558
               |||| |||  || |||| ||| | |  ||||| || ||||| || |||||  ||||
Db        2449 AGCGGAGTCAGTATCGAAGACTTATGGAAAATGAACAGTTTTGGGTTATTGGAGGTGTC 2508

Qy        2559 TCAGCCTATCTCTTC GGGTCTTCCA TTTC TTAAAAATGTTGGCTGGCATCGATACG 2618
               ||| |||  ||  || ||| |  | |||| |||||  |||| ||
Db        2509 TCAGCTTATTTCTTT GGTTTTTAA TTTC TCAAAATGCTAGCTGGTCTCGACACA 2568

Qy        2619 AACTTCACTGTCAAAAAAGAAGCTAAAAATGTAGAATTTGGGGAGCTATATATGGTC 2678
               || |||  |||  |  | | | ||  ||  | |||| ||||  || | |||
Db        2569 AATTTCACAGTCACATCGAAAACCGCAGATGATTTAGAATTCGGTGAGCTTTACATCGTC 2628

Qy        2679 AAGTGGACAACACTTTTGATTCCTCCAACCACACTTCTCATTATCAATATGTCGGG-- T 2735
               ,| |,|||,| || ,|||  |||  ||  ||||| || ||  ||    ||
Db        2629 AAATGGACAACTCTCTTGATCCCTCCGACTTCACTTCTTATAATTAACTTGGTCGGAGTC 2688

Qy        2736 TGTGCTGGATTCTCTGATGCACTCAACAAAGGATATGAAGCATGGGGGCCTCTCTTTGGC 2795
               ||||.|||| ||| || | |||. ||| |||||| |||| ||| . |||||
Db        2689 GTCGCTGGATTCTCTGATGCTCTTAACAAAGGTTATGAAGCTTGGGGACCTTTGTTTGGG 2748

Qy        2796 AAGGTGTTCTTTGCTTTCTGGGTGATTCTTCATCTCTATCCATTCCTTAAAGGTCTAATG 2855
               ||||| || | | || | |||||||| |||| |||||||||||  ,|.|||| |||
Db        2749 AAGGTCTTTTTCGCTTTTGGGTGATTCTTCATCTTTATCCATTCCTCAAAGGTCTTATG 2808

Qy        2856 GGTCGCCAAAACCTAACACCAACCATTGTTGTTCTCTGGTCAGTGCTGTTGGCCTCTGTC 2915
               ,| | |||,|  ,|||| || | ||| |||||||  | || .| ||||
Db        2809 GGAAGACAAAACAGAACACCTACTATTGTTATTCTCTGGTCTATCTTGCTTGCTTCTGTG 2868

Qy        2916 TTCTCTCTCGTTTGGGTCAAGATCAATCCATTCGTTAACAAAGTTGATAACACCTTGGTT 2975
               || || || |||'|||    | ||| ,| ||:|  ||||  |||  ,|  |  ||
Db        2869 TTTTCACTTGTTTGTGGTTCGTATCAACCCTTTCGTCTCCAAAACCGATACGACTTCCCTT 2928

Qy        2976 GCGGAGACCTGCATTTCCATTGATTGCTGA 3005
               |  |' |||  |||  || ||||||  |
Db        2929 TCTCTGAACTGTCTTTTGATCGATTGCTAA 2958
```

```
RESULT 7
AF150630
LOCUS       AF150630                3723 bp    mRNA    linear   PLN 04-FEB-2000
DEFINITION  Gossypium hirsutum cellulose synthase catalytic subunit (celA3)
            mRNA, complete cds.
ACCESSION   AF150630 AF200453
VERSION     AF150630.2  GI:6446576
KEYWORDS    .
SOURCE      Gossypium hirsutum (upland cotton)
  ORGANISM  Gossypium hirsutum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium.
REFERENCE   1  (bases 1 to 3723)
  AUTHORS   Laosinchai,W., Cui,X. and Brown,R.M. Jr.
  TITLE     A full length cDNA of cotton cellulose synthase has high homology
            with the Arabidopsis RSW1 gene and the cotton CelA1 gene (Accession
            No. AF200453) (PGR 00-002)
  JOURNAL   Plant Physiol. 122 (1), 291 (2000)
REFERENCE   2  (bases 1 to 3723)
  AUTHORS   Kimura,S., Laosinchai,W., Itoh,T., Cui,X. and Brown,R.M. Jr.
```

```
  TITLE     Immunogold Labeling of Rosette Terminal Cellulose Synthesizing
            Complexes in a Vascular Plant (Vigna angularis)
  JOURNAL   Unpublished
REFERENCE   3  (bases 1366 to 3470)
  AUTHORS   Laosinchai,W. and Brown,R.M. Jr.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-MAY-1999) Botany, University of Texas at Austin, W.
            24th Street, Austin, TX 78713, USA
REFERENCE   4  (bases 1 to 3723)
  AUTHORS   Laosinchai,W., Cui,X. and Brown,R.M. Jr.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-1999) Molecular Genetics and Microbiology,
            University of Texas at Austin, Austin, TX 78712, USA
  REMARK    Sequence update by submitter
COMMENT     On Nov 17, 1999 this sequence version replaced gi:5081778.
FEATURES             Location/Qualifiers
     source          1..3723
                     /organism="Gossypium hirsutum"
                     /mol_type="mRNA"
                     /cultivar="Texas marker-1"
                     /db_xref="taxon:3635"
                     /tissue_type="cotton fiber"
                     /dev_stage="24 days post anthesis"
     gene            1..3723
                     /gene="celA3"
     CDS             269..3472
                     /gene="celA3"
                     /note="celA1 homolog"
                     /codon_start=1
                     /product="cellulose synthase catalytic subunit"
                     /protein_id="AAD39534.2"
                     /db_xref="GI:6446577"
                     /translation="MESRSLIGGKPMKNLGGQTCQICGDNVGKNTDGDPFIACNICAF
                     PVCRPCYEYERKDGNQSCPQCKTRYKRWQKGSPAILDGRETGGDADDGASDFIYSENQE
                     QKQKLAERMQGWNAKYGRGEPVGADTYDKEISHNHTLLTSGQEVSGELSAASPERLS
                     MASPGVAGGKSSIRVVDEVREFGSSGLGNVAWKERVDGWKMKQEKNTVPMSTCQATSE
                     RGLGDIDASTDVLVDDSQLNGEARQPLSRKVSVSSSKINPYRMVIILRLVILCIFLHY
                     RITNPVPNAYALWLISVICEIWFAISWILDQFPKWLPVNRETYLDRLALRYDREGEPS
                     ELAAVDIFVSTVDPLKEPPLVTANTVLSILAVDYPVDKVSCYVSDDGAAMLTFEALSE
                     TSEFARKWVPFCKKYNIEPRAPEWYFAQKIDYLKDKVQTSFVKDRRAMKREYEEFKVR
                     INGLVAKAQKVPEEGWIMQDGTPWPGNNTRDHPGMIQVFLGQSGGLDAEGNELPRLVY
                     VSREKRPGFQHHKKAGAMNALVRVSAVLTNGAFLLNLDCDHYINNSKALREAMCFLMD
                     PNLGKQVCYVQFPQRFDGIDRNDRYANRNTVFFDINLRGLDGIQGPVYVGTGCVFNRT
                     ALYGYEPPLKPKHRKTGILSSLCGGSRKKSSKSSKKGSDKKKSGKHVDSTVPVFNLED
                     IEEGVEGAGFDDEKSLLMSQMSLEKRFGQSAVFVASTLMENGGVPQSATPETLLKEAI
                     HVISCGYEDKTDWGSEIGWIYGSVTEDILTGFKMHARGWRSIYCMPKRPAFKGSAPIN
                     LSDRLNQVLRWALGSVEILFSRHCPIWYGYSGRLKWLERFAYVNTTIYPVTAIPLLMY
                     CTLPAVCLLTNKFIIPQISNFASIWFISLFLSIFATGILEMKWNGVGIDQWWRNEQFW
                     VIGGVSAHLFAVFQGLLKVLAGVDTNFTVTSKAADDEAFAELYMFKWTTLLIPPTTL
                     LIINLVGVVAGISYVIN[illegible]
                     VVWSILLASIFSLLWVS[illegible]
BASE COUNT     1029 a    700 c    886 g    [illegible] t
ORIGIN

  Query Match           36.0%;  Score 2162.2;  DB 6;  Length 3723;
  Best Local Similarity   68.8%;  Pred. No. 1.2e-280;
  Matches 1695;  Conservative    0;  Mismatches  743;  Indels   27;  Gaps    6;

Qy        517 AACCGTCTGCAGAGGCTTCGGAGCCGCTTTCAATTGTTTATCAAATTCCACGCAACAAGC 576
Db        975 AACTGAATGATGAAGCTAGACAGCCTCTATCAAGGAAGGTTT[illegible]TTCCTCATCTAAAA 1034

Qy        577 TCACACCATACAGAGCAGTGATCATTAT[illegible]GACTGGTCATTC[illegible]GGCCTCTTCTTCCACT 636
Db       1035 TAAATCCTTATCGGATGGTCATCATACT[illegible]TTGTTATTC[illegible]GCATTTTCTTGCACT 1094

Qy        637 TCAGAATAACAAATCCTGTCGATAGTGCCTTTGGCCTGTGGCTTACTTCTGTCATATGTG 696
Db       1095 ACCGAATAACAAACCCTGTTCCAAATG[illegible]TATGTCTATGGT[illegible]ATATCTGTGATATGTG 1154

Qy        697 AGATCTGGTTTGCATTTTCTTCGGGT[illegible]GATAGTTCCCCAA[illegible]GGAATGCTGTCAATA 756
Db       1155 AGATTTGGTTTGCAATATCCTGGATA[illegible]GATAGTTCCCCAAATGGCTTC[illegible]TTAATC 1214

Qy        757 GAGAAACGTATATCGAAAGGCTGTCGGCAAGGTATGAAAGAGAGGGTGAGCCTTCTCAGC 816
Db       1215 GTGAGACGTATCTTGATAGGCTTGCCTTAAGATATGATCGAGAAGGAGAACCGTCGGAAT 1274

Qy        817 TTGCTGGTGTGGATTTTTTCGTGAGTACTGTTGATCCGCTGAAGGAACCGCCATTGATCA 876
Db       1275 TAGCTGCAGTTGACATATTTGTGAGTACTGTCGACCCATTAAAGGAGCCTCCACTTGTGA 1334

Qy        877 CTGCCAATACAGTCCTTTCCATCCTTGCTGTGGACTATCCCGTCGATAAAGTCTCCTGCT 936
Db       1335 CAGCCAACACTGTACTATCCATTCTTGCTGTTGACTACCCAGTTGACAAGGTCTCTTGCT 1394

Qy        937 ACGTGTCTGATGATGGTGCAGCTATGCTTTCATTTGAATCTCTTGTAGAAACAGCTGAGT 996
Db       1395 ATGTTTCTGATGATGGTGCTGCCATGTTAACCTTTGAAGCACTGTCTGAGACATCAGAGT 1454

Qy        997 TTGCAAGGAAGTGGGTTCCGTTCTGCAAAAAATTCTCAATTGAACCAAGAGCACCGGAGT 1056
Db       1455 TTGCCAGGAAATGGGTACCTTTCTGCAAGAAATATAACATTGAACCTCGGGCTCCAGAAT 1514

Qy       1057 TTTACTTCTCACAGAAAATTGATTACTTGAAAGACAAGGTTCAACCTTCTTTCGTGAAAG 1116
Db       1515 GGTATTTTGCACAGAAGATTGACTACCTGAAGGATAAAGTTCAAACATCTTTTGTTAAAG 1574

Qy       1117 AACGTAGAGCAATGAAAAGGGATTATGAAGAGTACAAAGTCCGAGTTAATGCCCTGGTAG 1176
Db       1575 ATCGTAGAGCTATGAAGAGAGAATATGAAGAATTTAAAGTTCGAATTAATGGGCTTGTCG 1634

Qy       1177 CAAAGGCTCAGAAAACACCTGAAGAAGGATGGACTATGCAAGATGGAACACCTTGGCCTG 1236
Db       1635 CAAAGGCACAGAAAGTTCCTGAAGAAGGATGGATTATGCAGGATGGTACACCATGGCCTG 1694

Qy       1237 GGAATAACACACGTGATCACCCTGGGCATGATTCAGGTCTTCCTTGGGAAATACTGGGAG 1296
Db       1695 GTAATAACACCAGAGACCATCC-AGGAATGATCCAGGTTTTCTTAGGCCAGAGTGGGGGC 1753

Qy       1297 CTCGTGACATTGAAGGAAATGAACTACCTCGTCTAGTATATGTCTCCAGGGAGAAGAGAC 1356
Db       1754 CT--TGATGCTGAGGGCAATGAGCTGCCAAGGTTAGTTTATGTTTCTCGTGAAAAGCGTC 1811

Qy       1357 CTGGCTACCAGCACCACAAAAAGGCTGGTGCAGAAAATGCTCTGGTGAGAGTGTCTGCAG 1416
Db       1812 CAGGCTTCCAACACCACAAAAAAGCTGGTGCTATGAATGCACTTGTTCGAGTGTCAGCAG 1871

Qy       1417 TACTCACAAATGCTCCCTACATCCTCAATGTTGATTGTGATCACTATGTAAACAATAGCA 1476
Db       1872 TCCTTACTAATGGAGCTTTCTTGTTAAATCTTGATTGTGATCACTACATAAATAATAGCA 1931

Qy       1477 AGGCTGTTCGAGAGGCAATGTGCATCCTGATGGACCCACAAGTAGGTCGAGATGTATGCT 1536
Db       1932 AGGCTTTAAGGGAAGCTATGTGTTTTCTGATGGATCCAAACCTTGGAAAGCAAGTTTGTT 1991

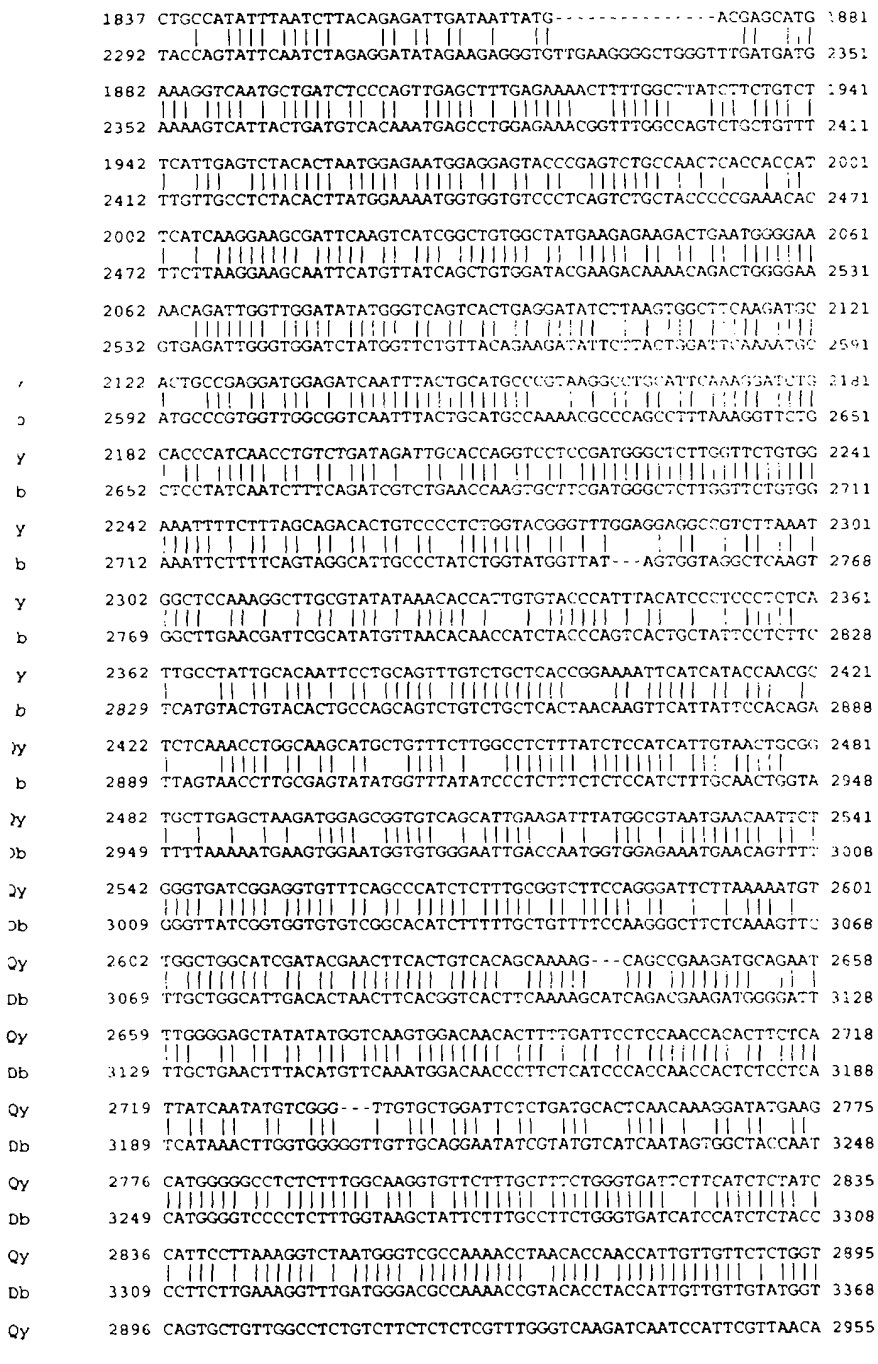
Qy       1537 ATGTGCAGTTCCCTCAGAGGTTTGATGGCATAGATAAGAGTGATCGCTACGCCAATCGTA 1596
Db       1992 ACGTTCAGTTTCCTCAAAGGTTTGATGGTATTGATCGGAATGATCGATATGCCAACAGAA 2051

Qy       1597 ACGTAGTTTTCTTTGATGTTAACATGAAAGGGTTGGATGGCATTCAAGGACCAGTATACG 1656
Db       2052 ATACTGTTTTCTTTGATATAAACTTGAGGGGTCTGGATGGCATTCAAGGGCCGGTTTATG 2111

Qy       1657 TAGGAACTGGTTGTGTTTTCAACAGGCAAGCACTTTACGGCTACGGGCCTCCTTCTATGC 1716
Db       2112 TCGGTACGGGATGTGTTTTCAATAGAACAGCATTGTATGGATATGAACCCCCCCTTAAGC 2171

Qy       1717 CCAGCTTACGCAAGAGAAAGGATTCTTCATCCTGCTTCTCATGTTGCTGCCCCCTCAAAGA 1776
Db       2172 CCAAGCATAGAAAAACAGGGATACTATCTTCATTGTGCGGAGGTTCTCGTAAGAAGAGTT 2231

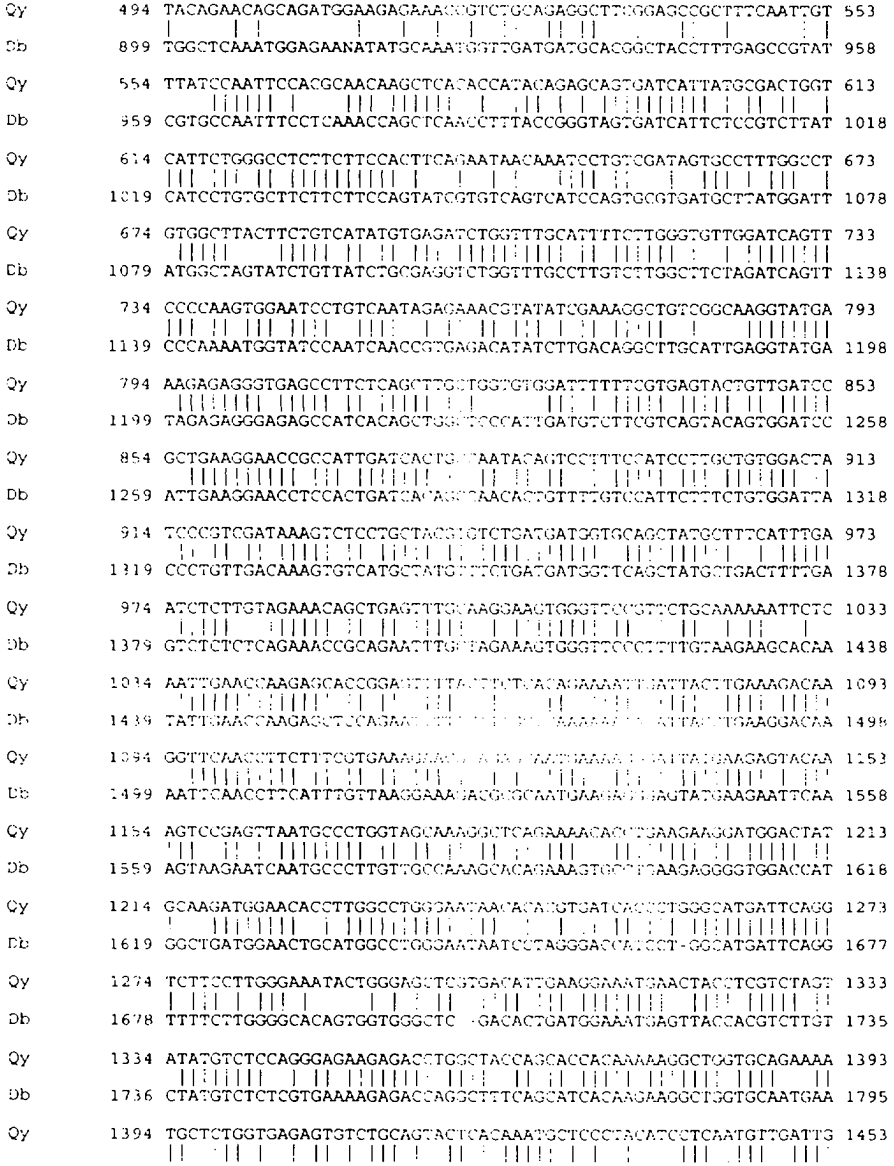
Qy       1777 AGAAGCCTGCTCAAGATCCAGCTGAGGTATACAGAGATGCAAAAAGAGAGGATCTCAATG 1836
Db       2232 CAAAATCAAGTAAAAAGGGATCAGACAAGAAGAAATCTGGCAAGCATGTTGATTCTACTG 2291
```

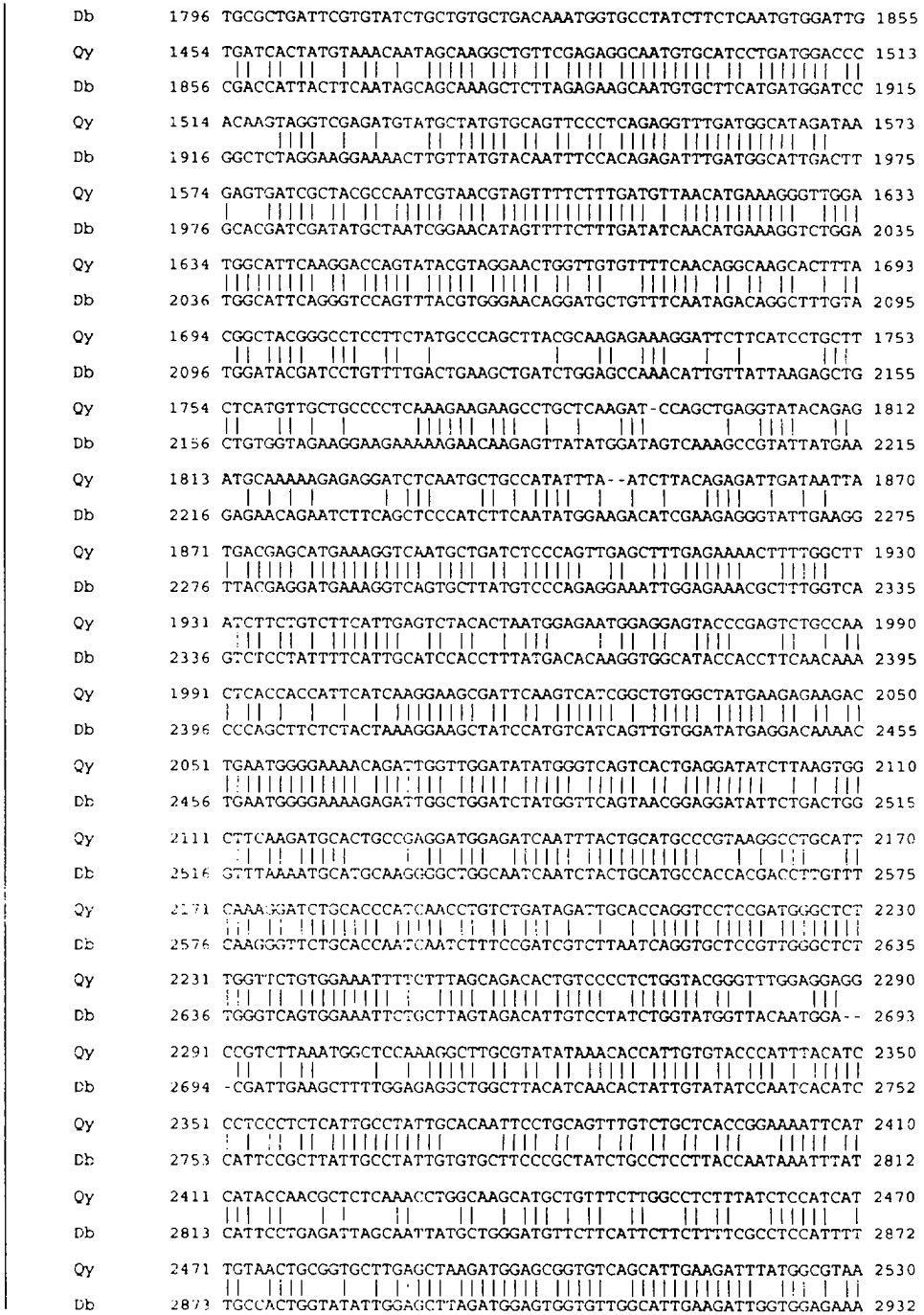


```
RESULT 8
AF200525
LOCUS       AF200525                [illegible] bp    mRNA    linear   PLN 31 AUG-2000
DEFINITION  Zea mays cellulose synthase 1 (CesA-1) mRNA, complete cds.
ACCESSION   AF200525
VERSION     AF200525.1  GI:[illegible]
KEYWORDS    .
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; [illegible]; Embryophyta; Tracheophyta;
            Spermatophyta; [illegible]; [illegible]; Poales; Poaceae; PACCAD
            clade; [illegible]; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to [illegible])
  AUTHORS   Holland,N., Holland,D., Whittier,T., Tabeau,N., Ehrhart,J.,
            Vanconostle,B., Hanzel,J., Haley,J., Bhagga,K.S.,
            Kubik,K., Kubik,C., Lai,J.S. and Ehrlich,S.D.
  TITLE     A comparative analysis of the plant cellulose synthase (CesA) gene
            family
  JOURNAL   Plant Physiol. 123 (4), 1313-1324 (2000)
  MEDLINE   20398328
   PUBMED   10938360
REFERENCE   2  (bases 1 to [illegible])
  AUTHORS   Bhagga,K.S. and Helentaris,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-NOV-1999) Agronomic Traits, Pioneer Hi-Bred
            International, Inc., 7250 NW 62nd Avenue, Johnston, IA 50131, USA
FEATURES             Location/Qualifiers
     source          1..[illegible]
                     /organism="Zea mays"
                     /mol_type="mRNA"
                     /db_xref="taxon:4577"
                     /note="identified and isolated from the maize genome
                     database cDNA collection of Pioneer Hi-Bred International,
                     Inc"
     gene            1..[illegible]
                     /gene="CesA 1"
     CDS             183..3410
                     /gene="CesA-1"
                     /note="CESA 1"
                     /codon_start=1
                     /product="cellulose synthase 1"
                     /protein_id="AAF89961.1"
                     /db_xref="GI:9622874"
                     /translation="MAANKGMVAGSHNRNEFVMIRHDGDVPGSAKPTKSANGQVCQIC
                     GDSVGVSATGDVFVACNECAFPVCRPCYEYERKEGNQCCPQCKTRYKRQKGSPRVHGD
                     EDEEDVDDLDNEFNYKQGSGKGPEWQLQGDDADLSSSARHEPHGRIPRLTSGQQISGR
                     IPDASPGRHSIRSPTSSYVDPSVPVPVRIVDPSKDLNSYGLNSVDWKERVESWRVKQD
                     KNMMQVCNKYPEARGGDMEDTSCNGEXMQMVDDARLPLSRIVPISSNQLNLYRVVIIL
                     RLIILCFFFQYRVSHPVRDAYGLWLVSVICEVWFALSWLLDQFPKWYPINRETYLDRL
                     ALRYDREGEPSQLAPIDVFVSTVDPLKEPPLITANTVLSILSVDYPVDKVSCYVSDDG
                     SAMLTFESLSETAEFARKWVPFCKKHNIEPRAPEFYFAQKIDYLKDKIQPSFVKERRA
                     MKREYEEFKVRINALVAKAQKVPEEGWTMADGTAWPGNNPRDHPGMIQVFLGHSGGLD
                     TDGNELPRLVYVSREKRPGFQHHKKAGAMNALIRVSAVLTNGAYLLNVDCDHYFNSSK
                     ALREAMCFMMDPALGRKTCYVQFPQRFDGIDLHDRYANRNIVFFDINMKGLDGIQGPV
                     YVGTGCCFNRQALYGYDPVLTEADLEPNIVIKSCCGRRKKKNKSYMDSQSRIMKRTES
                     SAPIFNMEDIEEGIEGYEDERSVLMSQRKLEKRFGQSPIFIASTFMTQGGIPPSTNPA
                     SLLKEAIHVISCGYEDKTEWGKEIGWIYGSVTEDILTGFKMHARGWQSIYCMPPRPCF
                     KGSAPINLSDRLNQVLRWALGSVEILLSRHCPIWYGYNGRLKLLERLAYINTIVYPIT
                     SIPLIAYCVLPAICLLTNKFIIPEISNYASMFFILLFASIFATGILELRWSGVGIEDW
                     WRNEQFWVIGGTSAHLFAVFQGLLKVLAGIDTNFTVTSKASDEDGDFAELYVFKWTSL
                     LIPPTTVLVINLVGMVAGISYAINSGYQSWGPLFGKLFFSIWVILHLYPFLKGLMGRQ
                     NRTPTIVIVWSILLASIFSLLWVKIDPFISPTQKAAALGQCGVNC"
BASE COUNT      944 a    804 c    965 g   1038 t      1 others
ORIGIN

  Query Match           35.5%;  Score 1148.8;  DB 8;  Length 3752;
```

Best Local Similarity 67.1%; Pred. No. 2.7e-277;
Matches 1734; Conservative 0; Mismatches 837; Indels 15; Gaps 7;





```
2531 TGAACAATTCTGGGTGATCGGAGGTGTTTCAGCCCATCTCTTTGCGGTCTTCCAGGGATT 2590
2933 TGAGCAGTTTTGGGTTATTGGTGGCACCTCTGCCCATCTCTTCGCAGTGTTCCAGGGTCT 2992

2591 CTTAAAAATGTTGGCTGGCATCGATACGAACTTCACTGTCACAGCAAAAGCA---GCCGA 2647
2993 GCTGAAAGTGTTGGCTGGGATTGATACCAACTTCACAGTTACCTCAAAGGCATCTGATGA 3052

2648 AGATGCAGAATTTGGGGAGCTATATATGGTCAAGTGGACAACACTTTTGATTCCTCCAAC 2707
3053 GGATGGCGACTTTGCTGAGCTATATGTGTTCAAGTGGACCAGTTTGCTCATTCCTCCGAC 3112

2708 CACACTTCTCATTATCAATATGTCGGG---TTGTGCTGGATTCTCTGATGCACTCAACAA 2764
3113 CACTGTTCTTGTCATTAACCTGGTCGGAATGGTGGCAGGAATTTCTTATGCCATTAACAG 3172

2765 AGGATATGAAGCATGGGGGCCTCTCTTTGGCAAGGTGTTCTTTGCTTTCTGGGTGATTCT 2824
3173 TGGCTACCAATCCTGGGGTCCGCTCTTTGGAAAGCTGTTCTTCTCGATCTGGGTGATCCT 3232

2825 TCATCTCTATCCATTCCTTAAAGGTCTAATGGGTCGCCAAAACCTAACACCAACCATTGT 2884
3233 CCATCTCTACCCCTTCCTCAAGGGTCTCATGGGAAGGCAGAACCGCACACCAACAATCGT 3292

2885 TGTTCTCTGGTCAGTGCTGTTGGCCTCTGTCTTCTCTCTCGTTTGGGTCAAGATCAATCC 2944
3293 CATTGTCTGGTCCATCCTTCTTGCATCTATCTTCTCCTTGCTGTGGGTGAAGATCGATCC 3352

2945 ATTCGT 2950
3353 TTTCAT 3358
```

```
SULT 9
652952
CUS       AX652952                3222 bp    DNA     linear   PAT 22-MAR-2003
LFINITION  Sequence 2822 from Patent WO03000898.
CESSION   AX652952
ERSION    AX652952.1  GI:29155766
LYWORDS   .
OURCE     Oryza sativa
 ORGANISM  Oryza sativa
           Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
           Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
           Ehrhartoideae; Oryzeae; Oryza.
EFERENCE  1
 AUTHORS   Chang,H.S., Chen,W., Cooper,B., Glazebrook,J., Goff,S.A., Hou,Y.M.,
           Katagiri,F., Quan,S., Tao,Y., Whitham,S., Xie,Z., Zhu,T. and Zou,G.
 TITLE     Plant genes involved in defense against pathogens
 JOURNAL   Patent: WO 03000898-A 2822 03-JAN-2003;
           Syngenta Participations AG (CH)
EATURES             Location/Qualifiers
     source         1..3222
                    /organism="Oryza sativa"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:4530"
ASE COUNT      805 a    711 c    805 g    901 t
RIGIN
```

```
 Query Match             35.5%;  Score 1146.2;  DB 6;  Length 3222;
 Best Local Similarity   68.6%;  Pred. No. 1.2e-276;
 Matches 1681;  Conservative    0;  Mismatches  753;  Indels   18;  Gaps    7;

y        509 GGAAGAGAAACCGTCTGCAGAGGCTTCGGAGCCGCTTTCAATTGTTTATCCAATTCCACG 568
o        726 GGACGCCTTACTGAATGATGAAACACGCCAACCTCTATCGAGAAAAGTGCCAATTTCGTC 785

y        569 CAACAAGCTCACACCATACAGAGCAGTGATCATTATGCGACTGGTCATTCTGGGCCTCTT 628
o        786 ATCCAGAATAAATCCGTACAGGATGGTTATTGTGCTCCGATTGATTGTCCTATGTATATT 845

y        629 CTTCCACTTCAGAATAACAAATCCTGTCGATAGTGCCTTTGGCCTGTGGCTTACTTCTGT 688
```

```
Db       846 CTTGCACTACCGTATCACAAATCCTGTGGTAATGCATACCCGCTATGGTTGCTCTCTGT 905

Qy       689 CATATGTGAGATCTGGTTTGCATTTTCTTGGTGTTGGATCAGTTCCCCAAGTGGAATCC 748
Db       906 CATATGTGAGATTTGGTTTGCTCTATCCTGGATTCTGGATCAGTTCCCCAAGTGGTCCCC 965

Qy       749 TGTCAATAGAGAAACGTATATCGAAAGGCTGTCGGCAAGGTATGAAAGAGAGGGTGAGCC 808
Db       966 AATCAACCGTGAAACTTACCTGGATAGGCTGGCTTTAAGGTATGACCGGGAAGGTGAACC 1025

Qy       809 TTCTCAGCTTGCTGGTGTGGATTTTTTCGTGAGTACTGTTGATCCGCTGAAGGAACCGCC 868
Db      1026 ATCTCAATTGGCTCCTGTTGACATTTTTGGCAGTACTGTGGATCCTATGAAAGAACCTCC 1085

Qy       869 ATTGATCACTGTCAATACAGTCTTTCTACTCTGCTGTGGACTATCCCGTCGATAAAGT 928
Db      1086 TCTTGTACTGTCAATACTTCCTCTCCTCGTGGATTACCCTGTTGATAAGGT 1145

Qy       929 GTCCTGTACCTTGGATGAAGCTAAATACTTTCATTTGAATCTCTTGTAGAAAC 988
Db      1146 ATCTTGCTATGTGTCTGATGATGGAGCTGAATGTTAACATTTGATGCGCTTGCTGAGAC 1205

Qy       989 AGCTGAGTTTGCAAGGAAGTGGGTTCCGTTCTGCAAAAAATTCTCAATTGAACCAAGAGC 1048
Db      1206 TTCAGAATTTGCTAGAAAATGGGTACCGTTCTGTAAAAAGTACAGCATAGAACCCAGAGC 1265

Qy       1049 ACCGGAGTTTTACTTCTCACAGAAAATTGATTACTTGAAAGACAAGGTTCAACCTTCTTT 1108
Db      1266 TCCAGAGTGGTACTTTGCACAGAAAATTGATTACTTGAAAGACAAAGTTCAGGCTTCTIT 1325

Qy       1109 CGTGAAAGAACGTAGAGCAATGAAAAGGGATTATGAAGAGTACAAAGTCCGAGTTAATGC 1168
Db      1326 TGTTAAAGATCGTCGTGCCATGAAGAGGGAATATGAAGAATTTAAAGTTCGTGTTAATGC 1385

Qy       1169 CCTGGTAGCAAAGGCTCAGAAAACACCTGAAGAAGGATGGACTATGCAAGATGGAACACC 1228
Db      1386 TCTCGTCGCCAAGGCACAGAAAGTTCCTGAGGAAGGATGGATTATGCAAGATGGCACACC 1445

Qy       1229 TTGGCCTGGGAATAACACACGTGATCACCTTGGGCATGATTCAGGTCTTCCTTGGGAAAT 1288
Db      1446 TTGGCCTGGTAACAATACCAGAGACCATCCTGGAATGATTCAGGTTTTCCTTGGTCA- 1502

Qy       1289 ACTGGGACCTCGTGACATTGAAGGAAATGAACTACCTCGTCTAGTATATGTCTCCAGGGA 1348
Db      1503 TAGTGGAGGCCTTGATACCGAAGGCAATGAGGTTCCTCGTTTAGTCTATGTGTCTCGTGA 1562

Qy       1349 GAAGAGACCTGGCTACCAGCACCACAAAAAGGCTGGTGCAGAAAATGCTCTGGTGAGAGT 1408
Db      1563 GAAACGTCCTGGATTCCAACACCATAAAAAGGCTGGTGCTATGAATGCACTTGTTCGTGT 1622

Qy       1409 GTCTGTAGTACTCACAAATGCTCCCTACATCCTCAATGTTGATTGTGATCACTATGTAAA 1468
Db      1623 ATCAGCTGTCCTTACTAATGGACAATACTTGTTGAATCTTGACTGTGATCACTACATCAA 1682

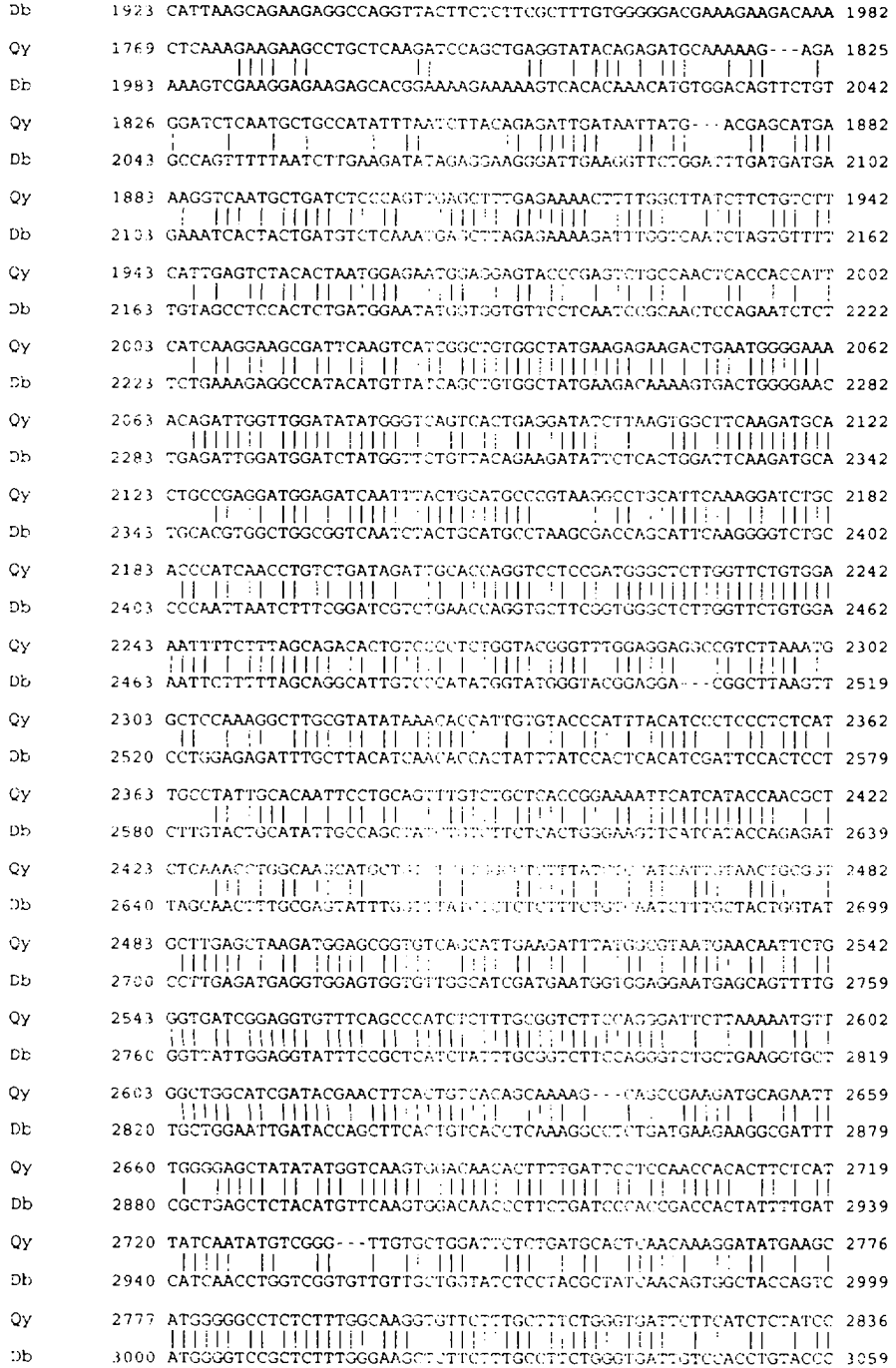
Qy       1469 CAATAGCAAGGCTGTTCGAGAGGCAATGTGCATCCTGATGGACCCACAAGTAGGTCGAGA 1528
Db      1683 CAATAGCAAAGCTTTCCGAGAGGCTATTGCTTCCTTATGGATCCAAACCTAGGAAGGCG 1742

Qy       1529 TGTATGTTATGTGCAGTTCCCTCAGAGGTTTGATGGCATAGATAAGAGTGATCGCTACGC 1588
Db      1743 TGTCTGTTATGTCCAATTCCCTCAGAGGTTTGACGGTATCGATAGAAATGATCGATATGC 1802

Qy       1589 CAATCGTAACGTAGTTTTCTTTGATGTTAACATGAAAGGGTTGGATGGCATTCAAGGACC 1648
Db      1803 AAACAGGAACACCGTGTTTTTCGATATTAATTTGAGAGGTCTCGATGGTCTCCAAGGACC 1862

Qy       1649 AGTATACGTAGGAACTGGTTGTGTTTTCAACAGGCAAGCACTTTACGGCTACGGGCCTCC 1708
Db      1863 AGTTTATGTGGGAACCGGTTGTGTGTTCAACAGAACAGCTCTTTATGGTTATGAACCCCC 1922

Qy       1709 TTCTATGCCCAGCTTACGCAAGAGAAAGGATTCTTCATCCTGCTTCTCATGTTGCTGCCC 1768
```



```
Qy       2837 ATTCCTTAAAGGTCTAATGGGTCGCCAAAACCTAACACCAACCATTGTTGTTCTCTGGTC 2896
Db       3060 CTTCCTTAAGGGTCTCATGGGTCGGCAGAACCGTACACCGACCATTGTTGTTGTTTGGGC 3119

Qy       2897 AGTGCTGTTGGCCTCTGTCTTCTCTCTCGTTTGGGTCAAGATCAATCCATTC 2948
Db       3120 CATCCTCCTTGCATCGATCTTCTCCTTGCTGTGGGTTCGTATCGATCCATTC 3171
```

```
RESULT 10
AF200529
LOCUS       AF200529               3676 bp    mRNA    linear   PLN 31-AUG-2000
DEFINITION  Zea mays cellulose synthase-5 (CesA-5) mRNA, complete cds.
ACCESSION   AF200529
VERSION     AF200529.1  GI:9622881
KEYWORDS    .
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 3676)
  AUTHORS   Holland,N., Holland,D., Helentjaris,T., Dhugga,K.S.,
            Xoconostle-Cazares,B. and Delmer,D.P.
  TITLE     A comparative analysis of the plant cellulose synthase (CesA) gene
            family
  JOURNAL   Plant Physiol. 123 (4), 1313-1324 (2000)
  MEDLINE   20398328
   PUBMED   10938350
REFERENCE   2  (bases 1 to 3676)
  AUTHORS   Dhugga,K.S. and Helentjaris,T.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-NOV-1999) Agronomic Traits, Pioneer Hi-Bred
            International, Inc., 7250 NW 62nd Avenue, Johnston, IA 50131, USA
FEATURES             Location/Qualifiers
     source          1. .3676
                     /organism="Zea mays"
                     /mol_type="mRNA"
                     /db_xref="taxon:4577"
                     /note="identified and isolated from the maize genome
                     database cDNA collection of Pioneer Hi-Bred International,
                     Inc"
     gene            1. .3676
                     /gene="CesA-5"
     CDS             254. .3484
                     /gene="CesA-5"
                     /note="CESA-5"
                     /codon_start=1
                     /product="cellulose synthase-5"
                     /protein_id="AAF89965.1"
                     /db_xref="GI:9622882"
                     /translation="MDGGDATNSGKHVAGQVCQICGDGVGTAADGDLFTACDVCGFPV
                     CRPCYEYERKDGTQACPQCKTKYKRHKGSPPVHGEENEDVDADDVSDYNYQASGNQDQ
                     KQKIAERMLTWRTNSRGSDIGLAKYDSGEIGHGKYDSGEIPRGYIPSLTHSQISGEIP
                     GASPDHMMSPVGNIGRRGHQFPYVNHSPNPSREFSGSLGNVAWKERVDGWKMKDKGAI
                     PMTNGTSIAPSEGRGVADIDASTDYNMEDALLNDETRQPLSRKVPIPSSRINPYRMVI
                     VLRLAVLCIFLRYRITHPVNNAYPLWLLSVICEIWFALSWILDQFPKWSPINRETYLD
                     RLALRYDREGEPSQLAPVDIFVSTVDPMKEPPLVTANTVLSILAVDYPVDKVSCYVSD
                     DGAAMLTFDALSETSEFARKWVPFCKKYNIEPXAPEWYFAQKIDYLKDKVQTSFVKER
                     RAMKREYEEFKVRINGLVAKAQKVPEEGWIMQDGTPWPGNNTRDHPGMIQVFLGHSGG
                     LDVEGNELPRLVYVSREKRPGFQHHKKAGAMNALVRVSAVLTNGQYMLNLDCDHYINN
                     SKALREAMCFLMDPNLGRNVCYVQFPQRFDGIDRNDRYANRNTVFFDINLRGLDGIQG
                     PVYVGTGCVFNRTALYGYEPPVKKKKPGFFSSLCGGRKKTSKSKKSSEKKKSHRHADS
                     SVPVFNLEDIEEGIEGSQFDDEKSLIMSQMSLEKRFGQSSVFVASTLMEYGGVPQSAT
                     PESLLKEAIHVISCGYEDKTDWGTEIGWIYGSVTEDILTGFKMHARGWRSIYCMPKRP
                     AFKGSAPINLSDRLNQVLRWALGSIEILFSRHCPIWYGYGGRLKFLERFAYINTTIYP
                     LTSIPLLLYCILPAVCLLTGKFIIPKISNLESVWFISLFISIFATGILEMRWSGVGID
                     EWWRNEQFWVIGGISAHLFAVFQGLLKVLAGIDTSFTVTSKATDEEGDFAELYMFKWT
                     TLLIPPTTILIINLVGVVAGISYAINSGYQSWGPLFGKLFFAFWVIVHLYPFLKGLMG
                     KQNRTPTIVVVWAILLASIFSLMWVRIDPFTTRVTGPDIAKCGINC"
BASE COUNT      907 a    826 c    951 g    991 t      1 others
ORIGIN
```

Query Match 35.2%; Score 1137.8; DB 8; Length 3676;
Best Local Similarity 67.8%; Pred. No. 1.6e-274;
Matches 1682; Conservative 0; Mismatches 784; Indels 15; Gaps 6;

```
 518 ACCGTCTGCAGAGGCTTCGGAGCCGCTTTCAATTGTTTATCCAATTCCACGCAACAAGCT 577
1000 ACTGAATGATGAAACTCGGCAACCTCTATCTAGAAAAGTGCCAATTCCTTCATCCAGAAT 1059

 578 CACACCATACAGAGCAGTGATCATTATGCGACTGGTCATTCTGGGCCTCTTCTTCCACTT 637
1060 AAATCCGTACAGAATGGTCATTGTGCTACGTTTGGCTGTTCTATGCATATTCTTGCGCTA 1119

 638 CAGAATAACAAATCCTGTCGATAGTGCCTTTGGCCTGTGGCTTACTTCTGTCATATGTGA 697
1120 CCGTATCACACATCCTGTGAACAATGCATATCCACTGTGGCTTTTATCCGTCATATGTGA 1179

 698 GATCTGGTTTGCATTTTCTTGGGTGTTGGATCAGTTCCCCAAGTGGAATCCTGTCAATAG 757
1180 GATCTGGTTTGCTTTGTCCTGGATTTTGGATCAGTTCCCAAAGTGGTCCCCAATCAAAGG 1239

 758 AGAAACGTATATCGAAAGGCTGTCGGCAAGGTATGAAAGAGAGGGTGAGCCTTCTCAGCT 817
1240 TGAAACATACCTTGATAGACTGGCTTTAAGGTATGACCGAGAAGGTGAACCATCTCAATT 1299

 818 TGCTGGTGTGGATTTTTTCGTGAGTACTGTTGATCCGCTGAAGGAACCGCCATTGATCAC 877
1300 AGCTCCTGTTGATATTTTTGTCAGTACTGTGGATCCAATGAAGGAGCCTCCTCTTGTCAC 1359

 878 TGCCAATACAGTCCTTTCCATCCTTGCTGTGGACTATCCCGTCGATAAAGTCTCCTGCTA 937
1360 TGCAAATACTGTGCTTTCCATCCTTGCTGTCGATTATCCGGTTGACAAGGTATCTTGCTA 1419

 938 CGTGTCTGATGATGGTGCAGCTATGCTTTCATTTGAATCTCTTGTAGAAACAGCTGAGTT 997
1420 TGTTTCGGATGATGGAGCTGCTATGCTGACTTTTGATGCTCTCTCTGAAACTTCAGAGTT 1479

 998 TGCAAGGAAGTGGGTTCCGTTCTGCAAAAAATTCTCAATTGAACCAAGAGCACCGGAGTT 1057
1480 TGCTAGAAAATGGGTTCCGTTCTGTAAGAAGTACAACATAGAGCCTANGGCCCCGGAATG 1539

1058 TTACTTCTCACAGAAAATTGATTACTTGAAAGACAAGGTTCAACCTTCTTTCGTGAAAGA 1117
1540 GTACTTTGCTCAGAAAATTGATTACTTGAAAGACAAAGTTCAAACCTCATTTGTGAAAGA 1599

1118 ACGTAGAGCAATGAAAAGGGATTATGAAGAGTACAAAGTCCGAGTTAATGCCCTGGTAGC 1177
1600 ACGCCGGGCCATGAAGAGAGAATATGAAGAATTCAAAGTTCGTATCAATGGTCTTGTAGC 1659

1178 AAAGGCTCAGAAAACACCTGAAGAAGGATGGACTATGCAAGATGGAACACCTTGGCCTGG 1237
1660 CAAGGCACAAAAAGTTCCCGAGGAGGGATGGATCATGCAAGATGGTACACCTTGGCCTGG 1719

1238 GAATAACACACGTGATCACCCTGGGCATGATTCAGGTCTTCCTTGGGAAATACTGGGAGC 1297
1720 GAACAATACTAGGGACCATCCT-GGAATGATTCAGGTTTTCCTGGGTCA--CAGTGGAGG 1776

1298 TCGTGACATTGAAGGAAATGAACTACCTCGTCTAGTATATGTCTCCAGGGAGAAGAGACC 1357
1777 GCTTGACGTTGAAGGCAATGAACTTCCTCGTTTGGTTTATGTGTCTCGTGAAAAACGTCC 1836

1358 TGGCTACCAGCACCACAAAAAGGCTGGTGCAGAAAATGCTCTGGTGAGAGTGTCTGCAGT 1417
1837 TGGATTCCAACATCACAAGAAGGCTGGTGCCATGAATGCACTTGTTCGTGTATCAGCTGT 1896

1418 ACTCACAAATGCTCCCTACATCCTCAATGTTGATTGTGATCACTATGTAAACAATAGCAA 1477
1897 CCTTACTAATGGGCAATACATGTTGAATCTTGATTGTGACCACTACATCAATAATAGCAA 1956

1478 GGCTGTTCGAGAGGCAATGTGCATCCTGATGGACCCACAAGTAGGTCGAGATGTATGCTA 1537
1957 GGCTCTTCGAGAAGCTATGTGCTTCCTTATGGACCCAAACCTAGGAAGGAATGTCTGTTA 2016
```

```
Qy   1538 TGTGCAGTT[illegible]AAGGTTTGA[illegible]ATAAGAGTGATCGCTACGCCAATCGTAA 1597
Db   2017 TGTCCAATT[illegible]AAGGTTTGA[illegible]ATAGGAATGACCGATATGCAAACAGGAA 2076

Qy   1598 CGTAGTTTT[illegible]ATGTTAACAT[illegible]GGATGGCATTCAAGGACCAGTATAGGT 1657
Db   2077 CACTGTGTTTTT[illegible]ATATTAACTT[illegible]TGACGGCATTCAAGGGCCAGTTTATGT 2136

Qy   1658 AGGAACTG[illegible]TTTTCAACA[illegible]TTACGGCTACGGGCCTCCTTCTATGCC 1717
Db   2137 GGGAACTG[illegible]TATAACAGAA[illegible]TATGGTTATGAGCCTCCAGTCAAGAA 2196

Qy   1718 CAGCTTAC[illegible]AAAGGATTC[illegible]TGCTTCTCATGTTGCTGCCCCTCAAA 1774
Db   2197 AAAAAAGC[illegible]TTTCTTCTT[illegible]GGGAAGGAAAAAGACGTCAAAATCTAA 2256

Qy   1775 GAAGAAC[illegible]AGAGATGCAAAAAGAGAGGATCTCAA 1834
Db   2257 [illegible]CAGACATTCTGTACCATTATTTAA 2316

Qy   1835 TGCTGC[illegible]TTACGACGAGCATGAAAGGTCAATGCT 1894
Db   2317 TCTCGAAGA[illegible]AGGAAGGGATT[illegible]TAGTTTGATGATGAGAAATCGCTGAT 2376

Qy   1895 GATCTCGCA[illegible]TTTATCTTCTGTCTTCATTGAGTCTAC 1954
Db   2377 TATGTCTCAAATGAGCTTGGAGAAGAGACTTC[illegible]AGTCCAGTGTTTTTGTAGGCTCTAC 2436

Qy   1955 ACTAATGGAGAAT[illegible]AGGAGTACC[illegible]AACTCACCACCATTCATCAAGGAAGC 2014
Db   2437 TCTGATGGAAATA[illegible]TGGTGTTC[illegible]AATTCGCAACTCCAGAGTCTCTTCTGAAAGAAGC 2496

Qy   2015 GATTCAAGTCATC[illegible]TGGCTATGAAGAGAAGACTGAATGGGGAAAACAGATTGGTTG 2074
Db   2497 TATTCATGT[illegible]ATCA[illegible]TGGCTATGAGGAGAAAACTGACTGGGGAACTGAGATTGGGTG 2556

Qy   2075 GATATATGGGTCAGTCACTGAGGATATCTTAA[illegible]GGCTTCAAGATGCACTGCCGAGGATG 2134
Db   2557 GATCTATGGTTCTGTTACAGAAGACATTCTCA[illegible]GGATTCAAGATGCATGCTCGAGGCTG 2616

Qy   2135 GAGATCAATTTACTGCATGCCCGTAAGGCCTGCATTCAAAGGATCTGCACCCATCAACCT 2194
Db   2617 GCGATCAATCTACTGCATGCCTAAGCGACCAGCTTTCAAGGGATCTGCTCCTATCAACCT 2676

Qy   2195 GTCTGATAGATTGCACCAGGTCCTCCGATC[illegible]TTGGTTCTGTGGAAATTTTCTTTAG 2254
Db   2677 TTCGGATCGTTTGAATCAAGTGCTTCGGTGGG[illegible]TTGGTTCCATTGAAATTCTTTTCAG 2736

Qy   2255 CAGACACTGTCCCCTCTGGTACGGGTTTGGAGGAGGCCGTCTTAAATGGCTCCAAAGGCT 2314
Db   2737 CAGGCATTGTCCCATATGGTATGGCTAT---GGAGGCCGGCTTAAATTCCTGGAGAGATT 2793

Qy   2315 TGCGTATATAAACACCATTGTGTACCCATTTACATCCCTCCCTCTCATTGCCTATTGCAC 2374
Db   2794 TGCTTATAT[illegible]AACACAACAATTTATCCACTCACATCAATCCCGCTCCTCCTGTACTGCAT 2853

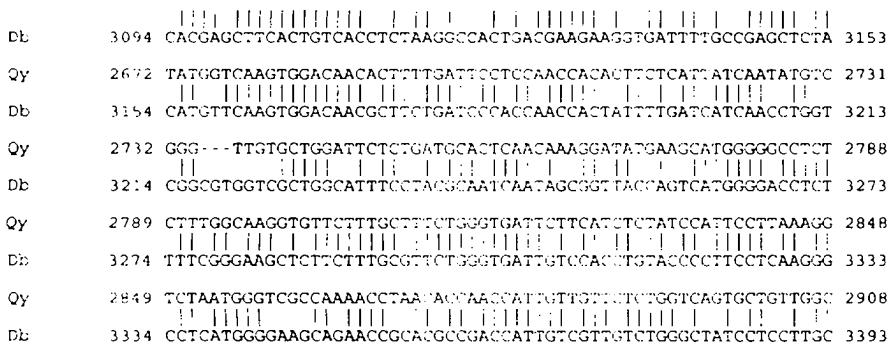
Qy   2375 AATTCCTGCAGTTTGTCTGCTCACCGGAAAATTCATCATACCAACGCTCTCAAACCTGGC 2434
Db   2854 ATTGCCAGCAGTTTGTCTTCTCACTGGGAAGTTCATCATCCCAAAGATTAGTAACCTAGA 2913

Qy   2435 AAGCATGCTGTTTCTTGGCCTCTTTATCTCCATCATTGTAACTGCGGTGCTTGAGCTAAG 2494
Db   2914 GAGTGTTTGGTTTATATCGCTCTTTATCTCAATCTTTGCCACTGGTATCCTTGAGATGAG 2973

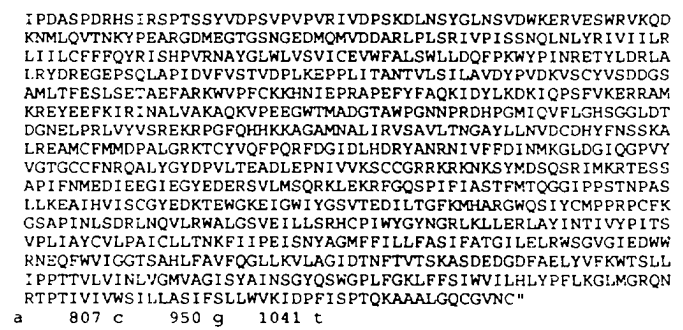
Qy   2495 ATGGAGCGGTGTCAGCATTGAAGATTTATGGCGTAATGAACAATTCTGGGTGATCGGAGG 2554
Db   2974 GTGGAGTGGTGTTGGCATTGATGAATGGTGGAGGAACGAGCAGTTCTGGGTCATTGGTGG 3033

Qy   2555 TGTTTCAGCCCATCTCTTTGCGGTCTTCCAGGGATTCTTAAAAATGTTGGCTGGCATCGA 2614
Db   3034 TATTTCTGCGCATTTATTTGCCGTCTTCCAGGGTCTCCTGAAGGTGCTTGCTGGTATCGA 3093

Qy   2615 TACGAACTTCACTGTCACAGCAAAAG---CAGCCGAAGATGCAGAATTTGGGGAGCTATA 2671
```

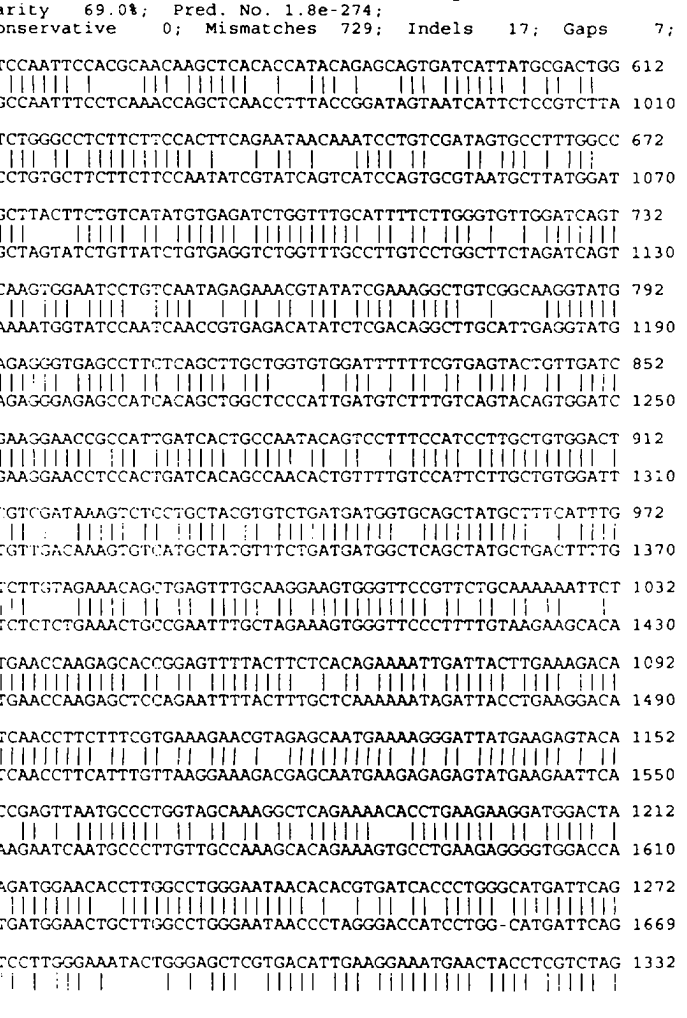


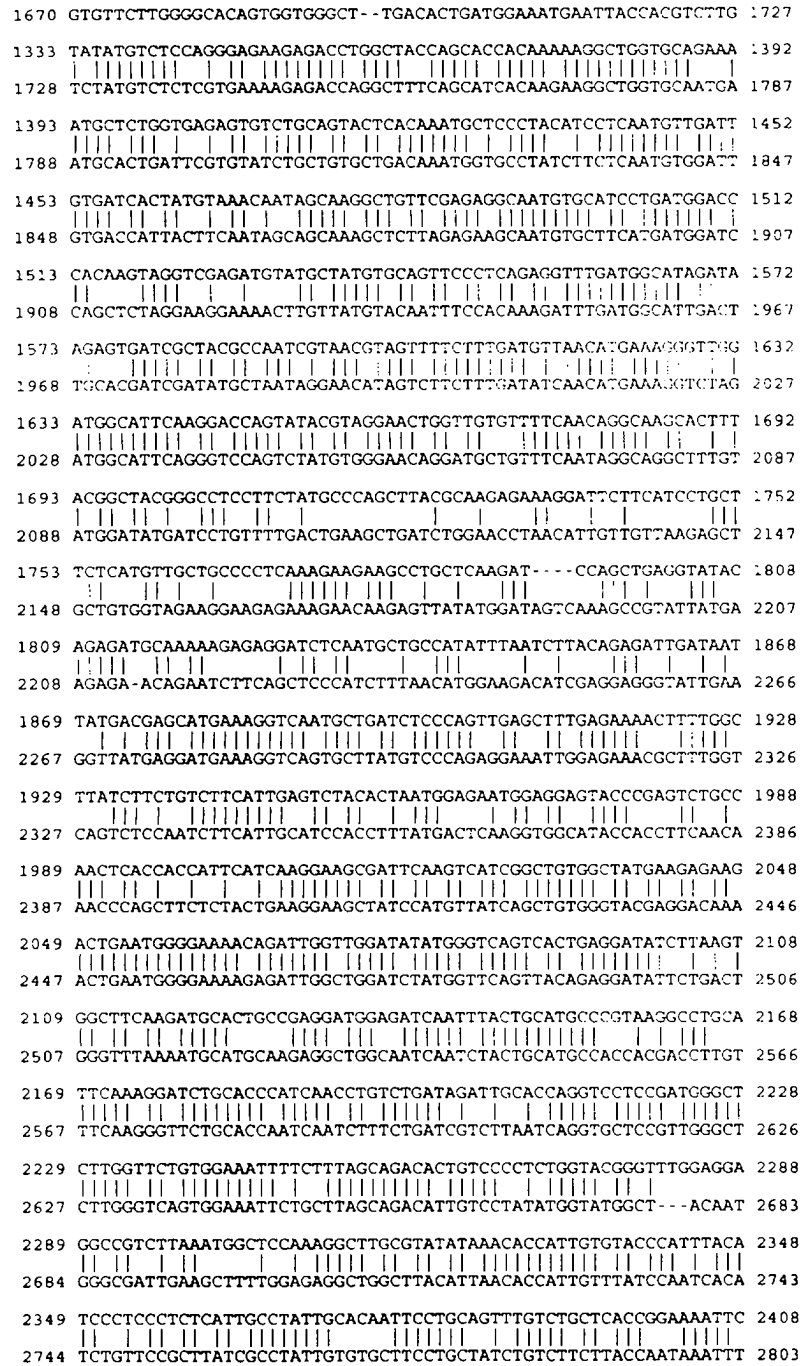
RESULT 11
AF200526
LOCUS AF200526 3725 bp mRNA linear PLN 31 AUG 2000
DEFINITION Zea mays cellulose synthase-2 (CesA 2) mRNA, complete cds.
ACCESSION AF200526
VERSION AF200526.1 GI:9622875
KEYWORDS .
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 3725)
AUTHORS Holland,N., Holland,D., Helentjaris,T., Dhugga,K.S., Xoconostle-Cazares,B. and Delmer,D.P.
TITLE A comparative analysis of the plant cellulose synthase (CesA) gene family
JOURNAL Plant Physiol. 123 (4), 1313-1324 (2000)
MEDLINE 20398428
PUBMED 10938358
REFERENCE 2 (bases 1 to 3725)
AUTHORS Dhugga,K.S. and Helentjaris,T.G.
TITLE Direct Submission
JOURNAL Submitted (01-NOV-1999) Agricultural Traits, Pioneer Hi-Bred International, Inc., 7250 NW 62nd Avenue, P.O. Box 552, Johnston, IA 50131, USA
FEATURES Location/Qualifiers
source 1..3725
/organism="Zea mays"
/mol_type="mRNA"
/db_xref="taxon:4577"
/note="identified and isolated from the maize genome database cDNA collection of Pioneer Hi-Bred International, Inc"
gene 1..3725
/gene="CesA-2"
CDS 179..3403
/gene="CesA-2"
/note="CESA-2"
/codon_start=1
/product="cellulose synthase-2"
/protein_id="AAF89962.1"
/db_xref="GI:9622876"
/translation="MAANKGMVAGSHNRNEFVMIHNQGDAPVPAKPTKSANQQVCQIC
GDTVGVSATGDVFVACNECAFPVCRPCYEYERKEGNQCCPQCKTRYKRQKGSPRVHGD
DEEEDVDDLDNEFNYKQGNGKGPEWQLQGDDADLNSSARHDFHHPIPRLTSGQQISGE

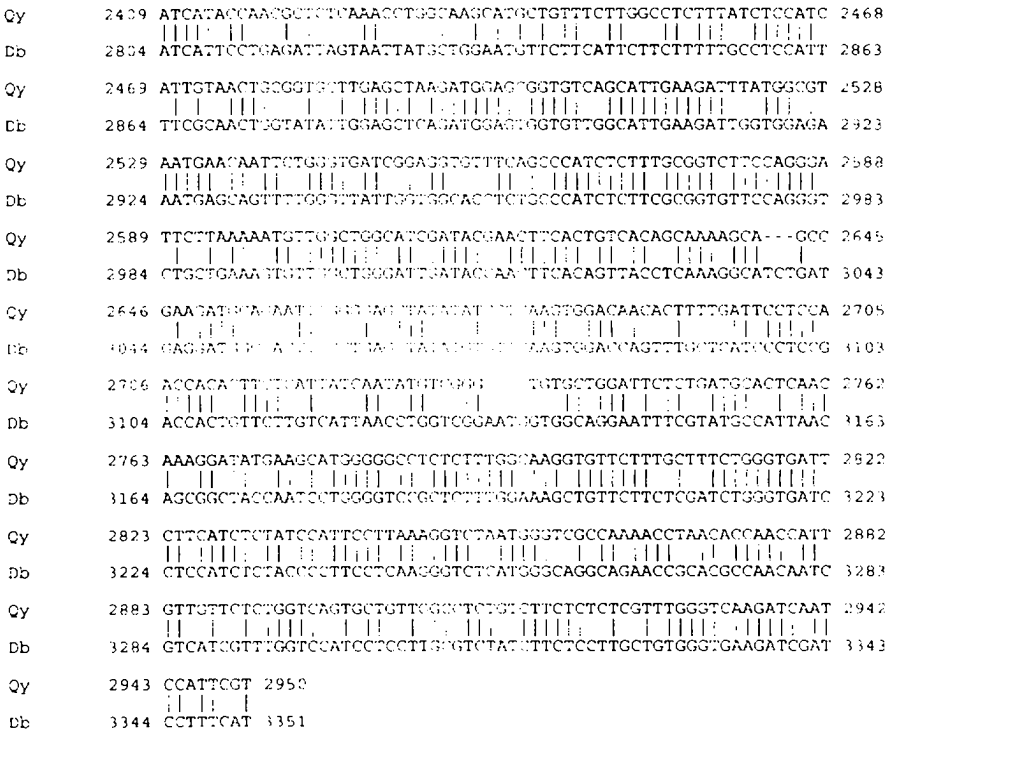


BASE COUNT 927 a 807 c 950 g 1041 t
ORIGIN

Query Match 35.2%; Score 1137.6; DB 8; Length 3725;
Best Local Similarity 69.0%; Pred. No. 1.8e-274;
Matches 1662; Conservative 0; Mismatches 729; Indels 17; Gaps 7;







RESULT 12
BT002335
LOCUS BT002335 3229 bp mRNA linear PLN 15-DEC-2002
DEFINITION Arabidopsis thaliana clone C104938 putative cellulose synthase catalytic subunit (At5g05170) mRNA, complete cds.
ACCESSION BT002335
VERSION BT002335.1 GI:26983831
KEYWORDS FLI_CDNA.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 3229)
AUTHORS Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M., Onodera,C.S., Quach,H.L., Tang,C.C., Toriumi,M., Wong,C., Wu,H.C., Yu,G., Yuan,S., Chen,H., Cheuk,R., Jones,T., Kim,C.J., Nguyen,M., Palm,C.C., Shinn,P., Southwick,A., Tripp,M.G., Wu,T., Davis,R.W., Ecker,J.R. and Theologis,A.
TITLE Arabidopsis Open Reading Frame (ORF) Clones
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 3229)
AUTHORS Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M., Onodera,C.S., Quach,H.L., Tang,C.C., Toriumi,M., Wong,C., Wu,H.C., Yu,G., Yuan,S., Chen,H., Cheuk,R., Jones,T., Kim,C.J., Nguyen,M., Palm,C.J., Shinn,P., Southwick,A., Tripp,M.G., Wu,T., Davis,R.W., Ecker,J.R. and Theologis,A.
TITLE Direct Submission
JOURNAL Submitted (15-DEC-2002) Plant Gene Expression Center, 800 Buchanan Street, Albany, CA 94710, USA
COMMENT Annotation based on July 2002 version of the Arabidopsis genome

```
          submitted to Genbank.
FEATURES             Location/Qualifiers
     source          1. .3229
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
                     /chromosome="5"
                     /clone="C104938"
                     /note="This clone is in pCNI 51."
                     ecotype: Columbia"
     gene            1. .3229
                     /gene="At5g05170"
     CDS             1. .3198
                     /gene="At5g05170"
                     /codon_start=1
                     /evidence=experimental
                     /product="putative cellulose synthase catalytic subunit"
                     /protein_id="AAN86168.1"
                     /db_xref="GI:26983832"
                     /translation="MESEGETAGKPMKNIVPQTCQICSDNVGKTVDGDRFVACDICSF
                     PVCRPCYEYERKDGNQSCPQCKTRYKRLKGSPAIPGDKDEDGLADEGTVEFNYPQKEK
                     ISERMLGWHLTRGKGEEMGFPQYDKEVSHNHLPRLTSRQDTSGEFSAASPERLSVSST
                     IAGGKRLPYSSDVNQSPNRRIVDPVGLGNVAWKERVDGWKMKQEKNTGPVSTQAASER
                     GGVDIDASTDILADEALLNDEARQPLSRKVSIPSSRINPYRMVIMLRLVILCLFLHYR
                     ITNPVPNAFALWLVSVICEIWFALSWILDQFPKWFPVNRETYLDRLALRYDREGEPSQ
                     LAAVDIFVSTVDPLKEPPLVTANTVLSILAVDYPVDKVSCYVSDDGAAMLSFESLAET
                     SEFARKWVPFCKKYSIEPRAPEWYFAAKIDYLKDKVQTSFVKDRRAMKREYEEFKIRI
                     NALVSKALKCPEEGWVMQDGTPWPGNNTRDHPGMIQVFLGQNGGLDAEGNELPRLVYV
                     SREKRPGFQHHKKAGAMNALVRVSAVLTNGPFILNLDCDHYINNSKALREAMCFLMDP
                     NLGKQVCYVQFPQRFDGIDKNDRYANRNTVFFDINLRGLDGIQGPVYVGTGCVFNRTA
                     LYGYEPPIKVKHKKPSLLSKLCGGSRKKNSKAKKESDKKKSGRHTDSTVPVFNLDDIE
                     EGVEGAGFDDEKALLMSQMSLEKRFGQSAVFVASTLMENGGVPPSATPENLLKEAIHV
                     ISCGYEDKSDWGMEIGWIYGSVTEDILTGFKMHARGWRSIYCMPKLPAFKGSAPINLS
                     DRLNQVLRWALGSVEIFFSRHCPIWYGYNGRLKFLERFAYVNTTIYPITSIPLLMYCT
                     LPAVCLFTNQFIIPQISNIASIWFLSLFLSIFATGILEMRWSGVGIDEWWRNEQFWVI
                     GGVSAHLFAVFQGILKVLAGIDTNFTVTSKASDEDGDFAELYLFKWTTLLIPPTTLLI
                     VNLVGVVAGVSYAINSGYQSWGPLFGKLFFAFWVIVHLYPFLKGLMGRQNRTPTIVVV
                     WSVLLASIFSLLWVRIDPFTSRVTGPDILECGINC"
     3'UTR           3199. .3229
                     /gene="At5g05170"
BASE COUNT      856 a    648 c    789 g    936 t
ORIGIN

  Query Match           35.1%;  Score 1135.4;  DB 8;  Length 3229;
  Best Local Similarity  68.1%;  Pred. No. 6.3e-274;
  Matches 1692;  Conservative    0;  Mismatches  761;  Indels   30;  Gaps    7;

Qy        519 CCGTCTGCAGAGGCTTCGGAGC[illegible]CAATTGTTATC[illegible]CAATTCTACGCAACAAGCTC 578
Db        706 CTGAATGACGAAGCGAGCCA[illegible]                                        765

Qy        579 ACACCATACAGAGCAGTGATCA[illegible]                                      638
Db        766 AATCCTTACAGAATGGTTATTATGCTGCGGCTTGTTATC[illegible]                    825

Qy        639 AGAATAACAAATCCTGTCGATAGTGCCTTTGGCCTGTGGC[illegible]                   698
Db        826 CGTATAACAAACCCAGTGCCAAATGCCTTTGCTCTATGGCTGGTCTCTGTGATATGTGAG 885

Qy        699 ATCTGGTTTGCATTTTCTTGGGTGTTGGATCAGTTCCCCAAGTGGAATCCTGTCAATAGA 758
Db        886 ATCTGGTTTGCCTTATCCTGGATTTTGGATCAGTTTCCCAAGTGGTTTCCTGTGAACCGT 945

Qy        759 GAAACGTATATCGAAAGGCTGTCGGCAAGGTATGAAAGAGAGGGTGAGCCTTCTCAGCTT 818
Db        946 GAAACCTACCTCGACAGGCTTGCTTTAAGATATGATCGTGAAGGTGAGCCATCACAGTTA 1005

Qy        819 GCTGGTGTGGATTTTTTCGTGAGTACTGTTGATCCGCTGAAGGAACCGCCATTGATCACT 878
Db       1006 GCTGCTGTTGACATTTTCGTGAGTACTGTTGACCCCTTGAAGGAGCCACCCCTTGTGACA 1065

Qy        879 GCCAATACAGTCCTTTCCATCCTTGCTGTGGACTATCCCGTCGATAAAGTCTCCTGCTAC 938
```

```
Db       1066 GCCAACACAGTGCTCTCTATTCTGGCTGTTGACTACCCAGTTGACAAGGTGTCCTGTTAT 1125

Qy        939 GTGTCTGATGATGGTGCAGCTATGCTTTCATTTGAATCTCTTGTAGAAACAGCTGAGTTT 998
Db       1126 GTTTCTGATGATGGTGCTGCTATGTTATCATTTGAATCACTTGCAGAAACATCAGAGTTT 1185

Qy        999 GCAAGGAAGTGGGTTCCGTTCTGCAAAAAATTCTCAATTGAACCAAGAGCACCGGAGTTT 1058
Db       1186 GCTCGTAAATGGGTACCATTTTGCAAGAAATATAGCATAGAGCCTCGTGCACCAGAATGG 1245

Qy       1059 TACTTCTCACAGAAAATTGATTACTTGAAAGACAAGGTTCAACCTTCTTTCGTGAAAGAA 1118
Db       1246 TACTTTGCTGCGAAAATAGATTACTTGAAGGATAAAGTTCAGACATCATTTGTCAAAGAT 1305

Qy       1119 CGTAGAGCAATGAAAAGGGATTATGAAGAGTACAAAGTCCGAGTTAATGCCCTGGTAGCA 1178
Db       1306 CGTAGAGCTATGAAGAGGGAATATGAGGAATTTAAAATCCGAATCAATGCACTTGTTTCC 1365

Qy       1179 AAGGCTCAGAAAACACCTGAAGAAGGATGGACTATGCAAGATGGAACACCTTGGCCTGGG 1238
Db       1366 AAAGCCCTAAAATGTCCTGAAGAAGGGTGGGTTATGCAAGATGGCACACCGTGGCCTGGA 1425

Qy       1239 AATAACACACGTGATCACCCTGGGCATGATTCAGGTCTTCCTTGGGAAATACTGGGAGCT 1298
Db       1426 AATAATACAAGGGACCATCC-AGGAATGATCCAGGTCTTCTTAGGGCAAAA--TGGTGGA 1482

Qy       1299 CGTGACATTGAAGGAAATGAACTACCTCGTCTAGTATATGTCTCCAGGGAGAAGAGACCT 1358
Db       1483 CTTGATGCAGAGGGCAATGAGCTCCCGCGTTTGGTATATGTTTCTCGAGAAAAGCGACCA 1542

Qy       1359 GGCTACCAGCACCACAAAAAGGCTGGTGCAGAAAATGCTCTGGTGAGAGTGTCTGCAGTA 1418
Db       1543 GGATTCCAGCACCACAAAAAGGCTGGTGCTATGAATGCACTGGTGAGAGTTTCAGCAGTT 1602

Qy       1419 CTCACAAATGCTCCCTACATCCTCAATGTTGATTGTGATCACTATGTAAACAATAGCAAG 1478
Db       1603 CTTACCAATGGACCTTTCATCTTGAATCTTGATTGTGATCATTACATAAATAACAGCAAA 1662

Qy       1479 GCTGTTCGAGAGGCAATGTGCATCCTGATGGACCCACAAGTAGGTCGAGATGTATGCTAT 1538
Db       1663 GCCTTAAGAGAAGCAATGTGCTTCCTGATGGACCCAAACCTCGGGAAGCAAGTTTGTTAT 1722

Qy       1539 GTGCAGTTCCCTCAGAGGTTTGATGGCATAGATAAGAGTGATCGCTACGCCAATCGTAAC 1598
Db       1723 GTTCAGTTCCCACAAAGATTTGATGGTATCGATAAGAACGATAGATATGCTAATCGTAAT 1782

Qy       1599 GTAGTTTTCTTTGATGTTAACATGAAAGGGTTGGATGGCATTCAAGGACCAGTATACGTA 1658
Db       1783 ACCGTGTTCTTTGATATTAACTTGAGAGGTTTAGATGGGATTCAAGGACCTGTATATGTC 1842

Qy       1659 GGAACTGGTTGTGTTTTCAACAGGCAAGCACTTTACGGCTACGGGCCTCCTTCTATGCCC 1718
Db       1843 GGAACTGGATGTGTTTTCAACAGAACAGCATTATACGGTTATGAACCTC---CAATAAAA 1899

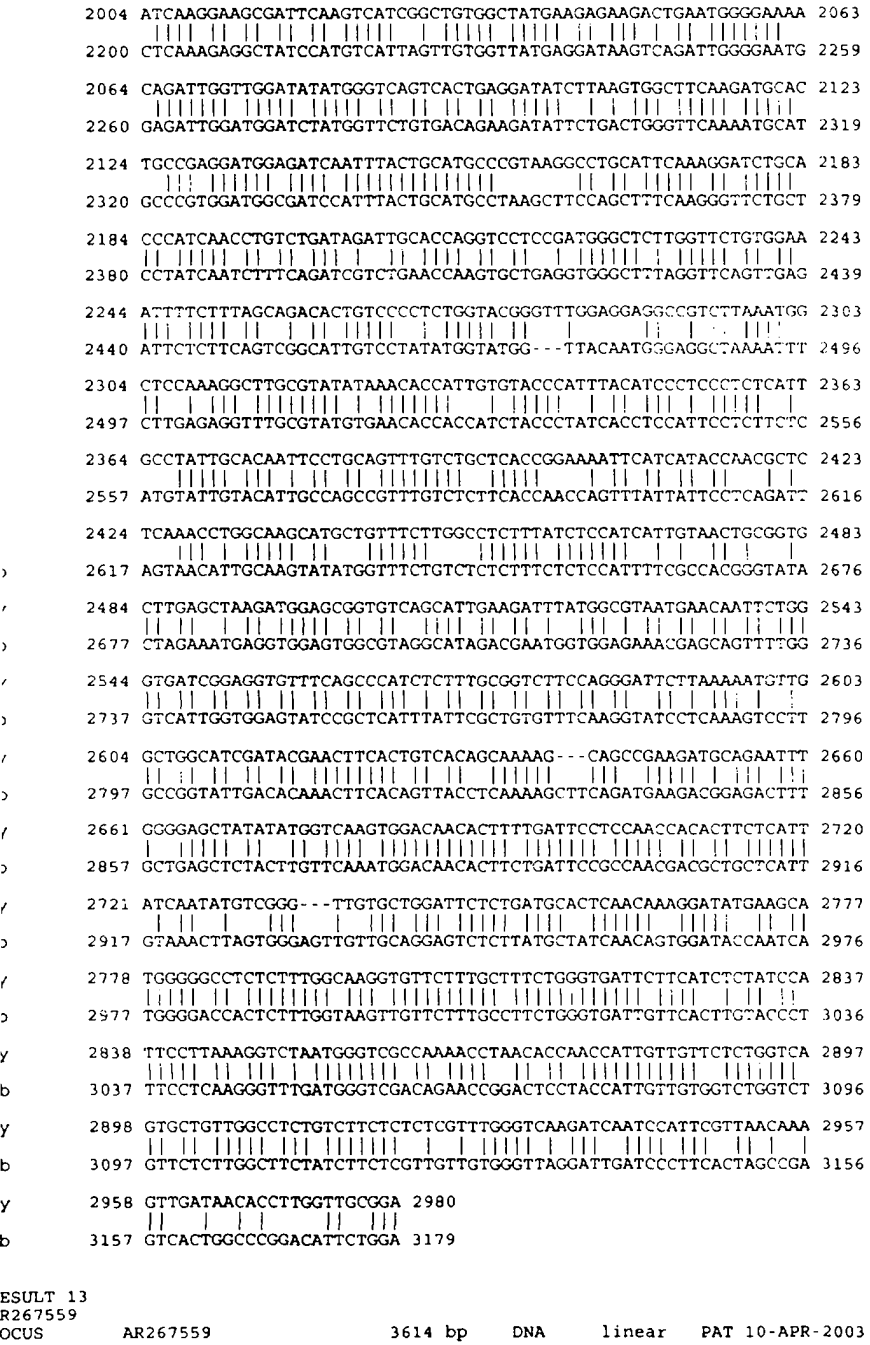
Qy       1719 AGCTTACGCAAGAGAAAGGATTCTTCATCCTGCTTCTCATGTTGCTGCCCCTCAAAGAAG 1778
Db       1900 GTAAAACACAAGAAGCCAAGTCTTTTATCTAAGCTCTGTGGTGGATCAAGAAAGAAGAAT 1959

Qy       1779 AAGCCTGCTCAAGATCCAGCTGAGGTATACAGAGATGCAAAAAGAGAGGATCTCAATGCT 1838
Db       1960 TCCAAAGCTAAGAAAGAGTCGGACAAAAAGAAATCAGGCAGGCATACTGACTCAACTGTT 2019

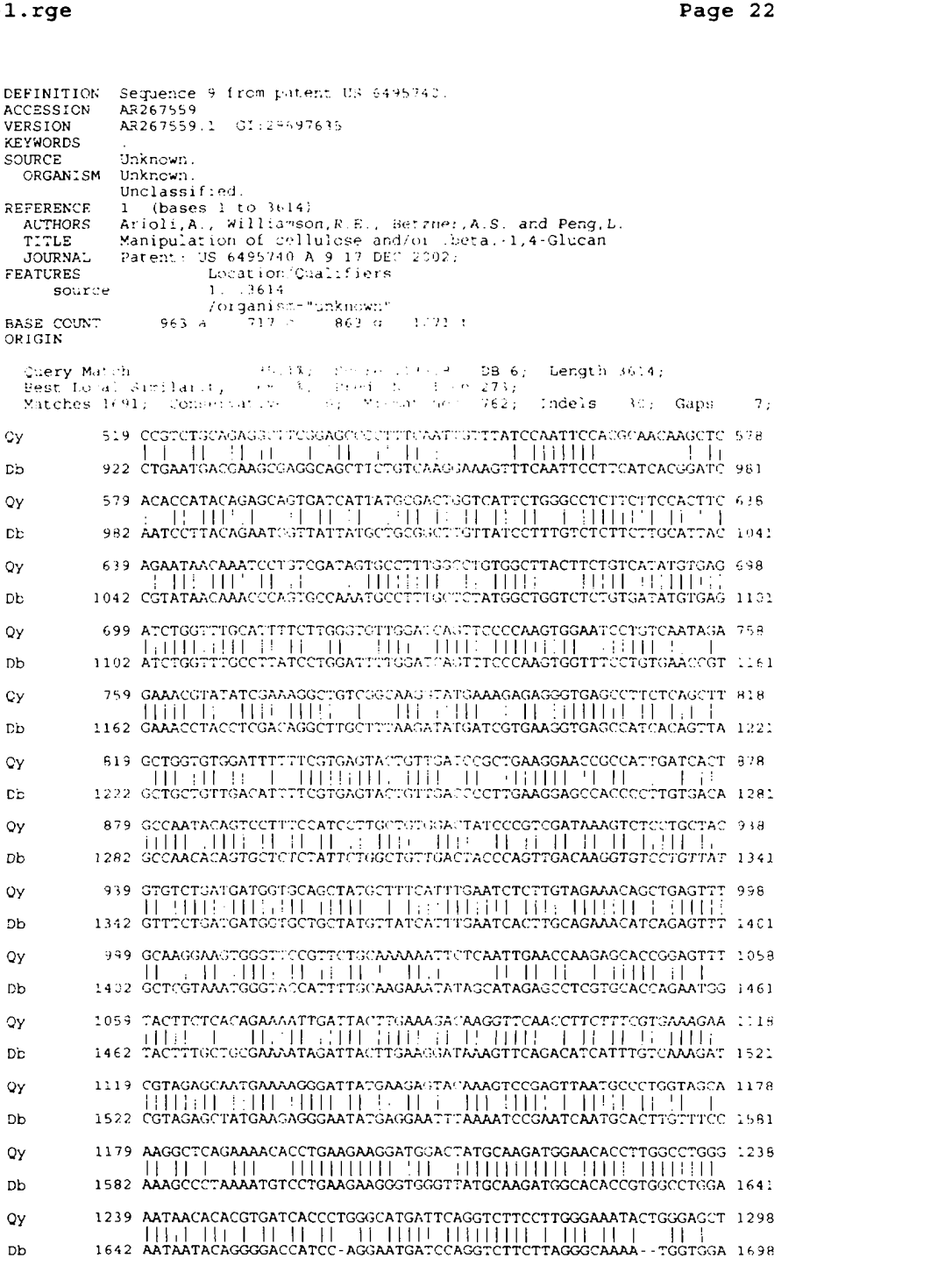
Qy       1839 GCCATATTTAATCTTACAGAGATTGATAA----------TTATGACGAGCATGAA 1883
Db       2020 CCTGTATTCAACCTCGATGACATAGAAGAGGGAGTTGAAGGTGCTGGTTTTGATGATGAA 2079

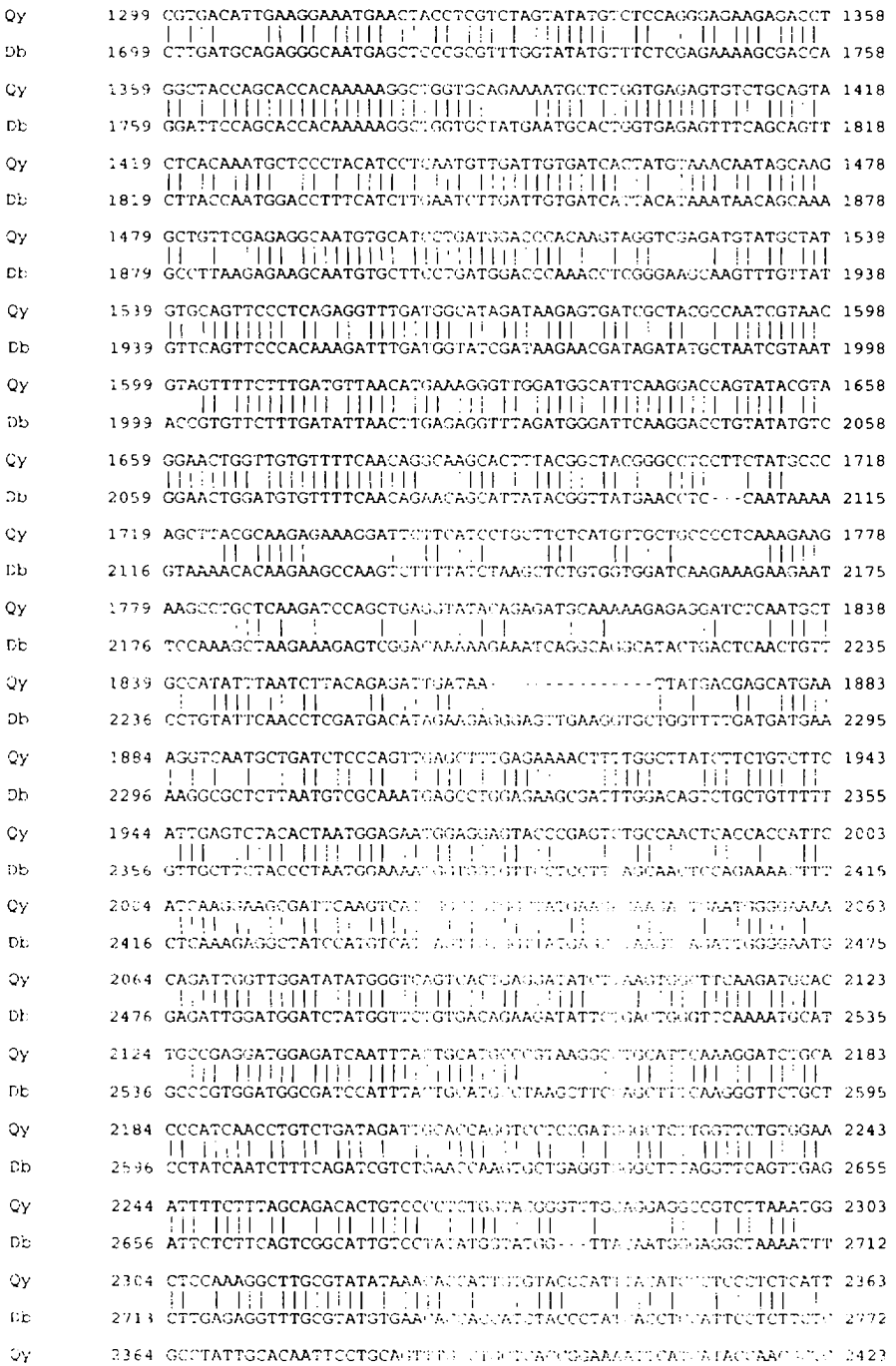
Qy       1884 AGGTCAATGCTGATCTCCCAGTTGAGCTTTGAGAAAACTTTTGGCTTATCTTCTGTCTTC 1943
Db       2080 AAGGCGCTCTTAATGTCGCAAATGAGCCTGGAGAAGCGATTTGGACAGTCTGCTGTTTTT 2139

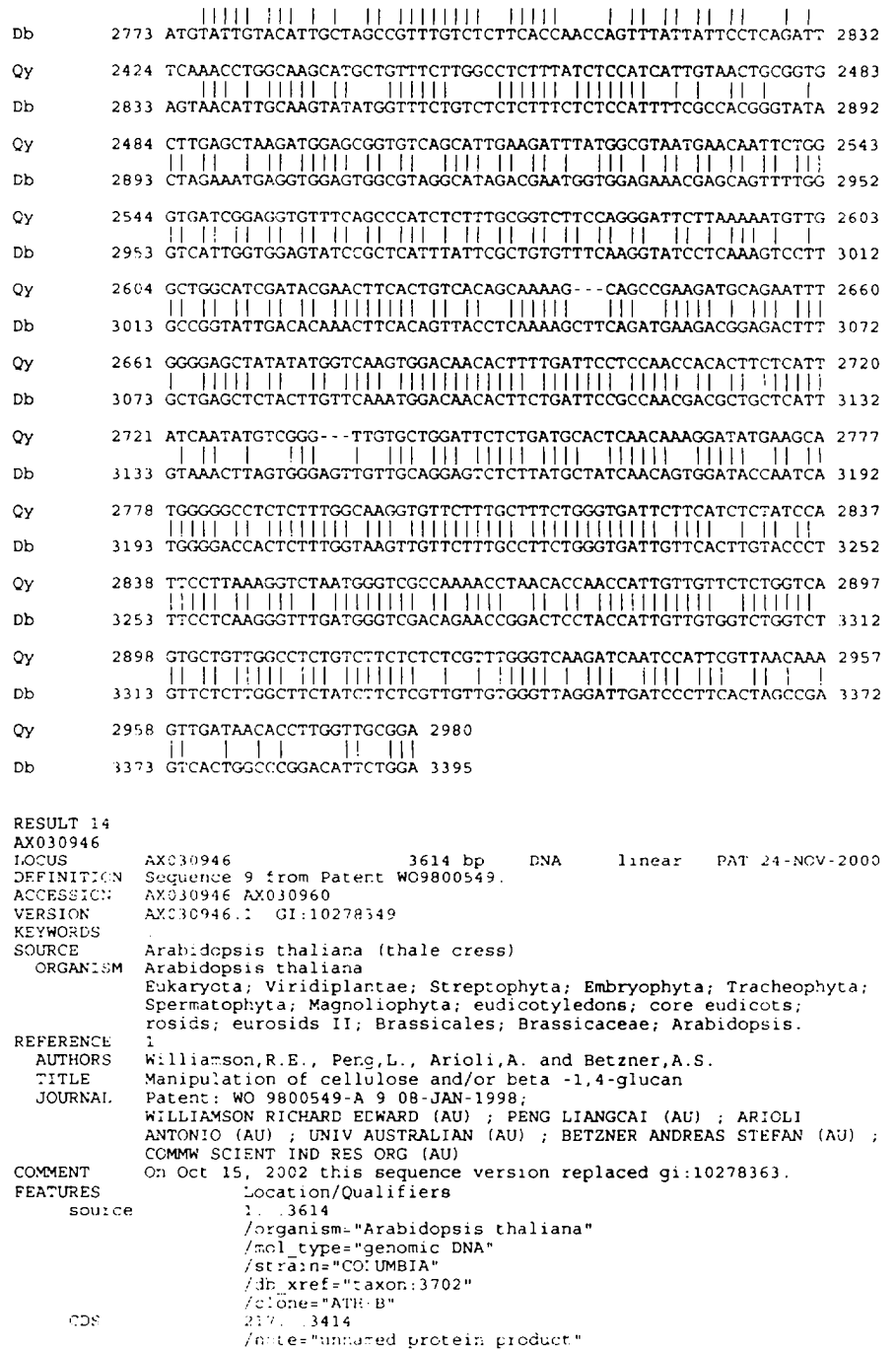
Qy       1944 ATTGAGTCTACACTAATGGAGAATGGAGGAGTACCCGAGTCTGCCAACTCACCACCATTC 2003
Db       2140 GTTGCTTCTACCCTAATGGAAAATGGTGGTGTTCCTCCTTCAGCAACTCCAGAAAACCTT 2199
```



ESULT 13
R267559
OCUS AR267559 3614 bp DNA linear PAT 10-APR-2003







```
RESULT 14
AX030946
LOCUS       AX030946               3614 bp    DNA     linear   PAT 24-NOV-2000
DEFINITION  Sequence 9 from Patent WO9800549.
ACCESSION   AX030946 AX030960
VERSION     AX030946.1  GI:10278349
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Williamson,R.E., Peng,L., Arioli,A. and Betzner,A.S.
  TITLE     Manipulation of cellulose and/or beta -1,4-glucan
  JOURNAL   Patent: WO 9800549-A 9 08-JAN-1998;
            WILLIAMSON RICHARD EDWARD (AU) ; PENG LIANGCAI (AU) ; ARIOLI
            ANTONIO (AU) ; UNIV AUSTRALIAN (AU) ; BETZNER ANDREAS STEFAN (AU) ;
            COMMW SCIENT IND RES ORG (AU)
COMMENT     On Oct 15, 2002 this sequence version replaced gi:10278363.
FEATURES             Location/Qualifiers
     source          1..3614
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /strain="COLUMBIA"
                     /db_xref="taxon:3702"
                     /clone="ATR1B"
     CDS             217..3414
                     /note="unnamed protein product"
```

```
                /codon_start=1
                /protein_id="CAC09627.1"
                /db_xref="GI:10278350"
                /db_xref="REMTREMBL:CAC09627"
                /translation="MESEGETAGKPMKNIVPQTCQICSDNVGKTVDGDRFVACDICSF
                PVCRPCYEYERKDGNQSCPQCKTRYKRLKGSPAIPGDKDEDGLADEGTVEFNYPQKEK
                ISERMLGWHLTRGKGEEMGEPQYDKEVSHNHLPRLTSRQDTSGEFSAASPERLSVSST
                IAGGKRLPYSSDVNQSPNRRIVDPVGLGNVAWKERVDGWKMKQEKNTGPVSTQAASER
                GGVDIDASTDILADEALLNDEARQLLSRKVSIPSSRINPYRMVIMLRLVILCLFLHYR
                ITNPVPNAFALWLVSVICEIWFALSWILDQFPKWFPVNRETYLDRLALRYDREGEPSQ
                LAAVDIFVSTVDPLKEPPLVTANTVLSILAVDYPVDKVSCYVSDDGAAMLSFESLAET
                SEFARKWVPFCKKYSIEPRAPEWYFAAKIDYLKDKVQTSFVKDRRAMKREYEEFKIRI
                NALVSKALKCPEEGWVMQDGTPWPGNNTGDHPGMIQVFLGQNGGLDAEGNELPRLVYV
                SREKRPGFQHHKKAGAMNALVRVSAVLTNGPFILNLDCDHYINNSKALREAMCFLMDP
                NLGKQVCYVQFPQRFDGIDKNDRYANRNTVFFDINLRGLDGIQGPVYVGTGCVFNRTA
                LYGYEPPIKVKHKKPSLLSKLCGGSRKKNSKAKKESDKKKSGRHTDSTVPVFNLDDIE
                EGVEGAGFDDEKALLMSQMSLEKRFGQSAVFVASTLMENGGVPPSATPENFLKEAIHV
                ISCGYEDKSDWGMEIGWIYGSVTEDILTGFKMHARGWRSIYCMPKLPAFKGSAPINLS
                DRLNQVLRWALGSVEILFSRHCPIWYGYNGRLKFLERFAYVNTTIYPITSIPLIMYCT
                LLAVCLFTNQFIIPQISNIASIWFLSLFLSIFATGILEMRWSGVGIDEWWRNEQFWVI
                GGVSAHLFAVFQGILKVLAGIDTNFTVTSKASDEDGDFAELYLFKWTTLLIPPITLLI
                VNLVGVVAGVSYAINSGYQSWGPLFGKLFFAFWVIVHLYPFLKGLMGRQNRTPTIVVV
                WSVLLASIFSLLWVRIDPFTSRVTGPDILECGINC"
SE COUNT      963 a    717 c    863 g   1071 t
IGIN

Query Match            35.1%;  Score 1133.8;  DB 6;  Length 3614;
Best Local Similarity  68.1%;  Pred. No. 1.6e-273;
Matches 1691;  Conservative    0;  Mismatches  762;  Indels   30;  Gaps    7;

Qy        519 CCGTCTGCAGAGGCTTCGGAGCCGCTTTCAATTGTTTATCCAATTCCACGCAACAAGCTC 578
Db        922 CTGAATGACGAAGCGAGGCAGCTTCTGTCAAGGAAAGTTTCAATTCCTTCATCACGGATC 981

Qy        579 ACACCATACAGAGCAGTGATCATTATGCGACTGGTCATTCTGGGCCTCTTCTTCCACTTC 638
Db        982 AATCCTTACAGAATGGTTATTATGCTGCGGCTTGTTATCCTTTGTCTCTTCTTGCATTAC 1041

Qy        639 AGAATAACAAATCCTGTCGATAGTGCCTTTGGCCTGTGGCTTACTTCTGTCATATGTGAG 698
Db       1042 CGTATAACAAACCCAGTGCCAAATGCCTTTGCTCTATGGCTGGTCTCTGTGATATGTGAG 1101

Qy        699 ATCTGGTTTGCATTTTCTTGGGTGTTGGATCAGTTCCCCAAGTGGAATCCTGTCAATAGA 758
Db       1102 ATCTGGTTTGCCTTATCCTGGATTTTGGATCAGTTTCCCAAGTGGTTTCCTGTGAACCGT 1161

Qy        759 GAAACGTATATCGAAAGGCTGTCGGCAAGGTATGAAAGAGAGGGTGAGCCTTCTCAGCTT 818
Db       1162 GAAACCTACCTCGACAGGCTTGCTTTAAGATATGATCGTGAAGGTGAGCCATCACAGTTA 1221

Qy        819 GCTGGTGTGGATTTTTTCGTGAGTACTGTTGATCCGCTGAAGGAACCGCCATTGATCACT 878
Db       1222 GCTGCTGTTGACATTTTCGTGAGTACTGTTGACCCCTTGAAGGAGCCACCCCTTGTGACA 1281

Qy        879 GCCAATACAGTCCTTTCCATCCTTGCTGTGGACTATCCCGTCGATAAAGTCTCCTGCTAC 938
Db       1282 GCCAACACAGTGCTCTCTATTCTGGCTGTTGACTACCCAGTTGACAAGGTGTCCTGTTAT 1341

Qy        939 GTGTCTGATGATGGTGCAGCTATGCTTTCATTTGAATCTCTTGTAGAAACAGCTGAGTTT 998
Db       1342 GTTTCTGATGATGGTGCTGCTATGTTATCATTTGAATCACTTGCAGAAACATCAGAGTTT 1401

Qy        999 GCAAGGAAGTGGGTTCCGTTCTGCAAAAAATTCTCAATTGAACCAAGAGCACCGGAGTTT 1058
Db       1402 GCTCGTAAATGGGTACCATTTTGCAAGAAATATAGCATAGAGCCTCGTGCACCAGAATGG 1461

Qy       1059 TACTTCTCACAGAAAATTGATTACTTGAAAGACAAGGTTCAACCTTCTTTCGTGAAAGAA 1118
Db       1462 TACTTTGCTGCGAAAATAGATTACTTGAAGGATAAAGTTCAGACATCATTTGTCAAAGAT 1521

Qy       1119 CGTAGAGCAATGAAAAGGGATTATGAAGAGTACAAAGTCCGAGTTAATGCCCTGGTAGCA 1178
Db       1522 CGTAGAGCTATGAAGAGGGAATATGAGGAATTTAAAATCCGAATCAATGCACTTGTTTCC 1581

Qy       1179 AAGGCTCAGAAAA[illegible]TGAAGAAGTATGGACTATGCAAGATGGAACACCTTGGCCTGGG 1238
Db       1582 AAAGCCCTAAAATGT[illegible]GAAGAAGTGGGTTATGCAAGATGGCACACCGTGGCCTGGA 1641

Qy       1239 AATAACACAC[illegible]ATCCTGGGCATGATTCAGGTCTTCCTTGGGAAATACTGGGAGCT 1298
Db       1642 AATAATACAC[illegible]ATCC-AGGAATGATCCAGGTCTTCTTAGGGCAAAA--TGGTGGA 1698

Qy       1299 CGTGACATT[illegible]TAGTATATGTCTCCAGGGAGAAGAGACCT 1358
Db       1699 CTTGATGC[illegible]TATGAGTCCGCGTTTGTATATGTTTCTCGAGAAAAGCGACCA 1758

Qy       1359 GGCTACCA[illegible]AAAAAGGCTGGTGCAGAAAATGCTCTGGTGAGAGTGTCTGCAGTA 1418
Db       1759 GGATTCCA[illegible]AAAAAGGCTGGTGCTATGAATGCACTGGTGAGAGTTTCAGCAGTT 1818

Qy       1419 CTCACAAAT[illegible]TGATCACTATGTAAACAATAGCAAG 1478
Db       1819 CTTACAAA[illegible]TGATCATTACATAAATAACAGCAAA 1878

Qy       1479 GCTGTT[illegible]ACAAGTAGGTCGAGATGTATGCTAT 1538
Db       1879 GCCTTAAGA[illegible]AATGTGCTTCCTGATGGACCCAAACCTCGGGAAGCAAGTTTGTTAT 1938

Qy       1539 GTGCAGTTCCCTCAGAGGTTTGATGCCATAGATAAGAGTGATCGCTACGCCAATCGTAAC 1598
Db       1939 GTTCAGTTCCCACAAAGATTTGATGGTATTGATAAGAACGATAGATATGCTAATCGTAAT 1998

Qy       1599 GTAGTTTTCTTTGACGTTAACATGAAAGG[illegible]GATGGCATTCAAGGACCAGTATACGTA 1658
Db       1999 ACCGTGTTCTTTGATATTAACTTGAAGGGTTTAGATGGGATTCAAGGACCTGTATATGTC 2058

Qy       1659 GGAACTGGTTGTGTTTTCAACAGGCAAGCACTTTACGGCTACGGGCCTCCTTCTATGCCC 1718
Db       2059 GGAACTGGATGTGTTTTCAACAGAACAGCATTATACGGTTATGAACCTC  CAATAAAA 2115

Qy       1719 AGCTTACGCAAGAGAAAGGATTCTTCATCTTGCTTCTCATGTTGCTGCCCCTCAAAGAAG 1778
Db       2116 GTAAAACACAAGAAGCCAAGTCTTTTATCTAAGCTCTGTGGTGGATCAAGAAAGAAGAAT 2175

Qy       1779 AAGCCTGCTCAAGATCCAGCTGACGTATACAGAGATGCAAAAAGAGAGGATCTCAATGCT 1838
Db       2176 TCCAAAGCTAAGAAA[illegible]AGTCGGACAAAAAGAAATCAGGCAGGCATACTGACTCAACTGTT 2235

Qy       1839 GCCATATTTAATCTTACAGAGATTGATAA ----------TTATGACGAGCATGAA 1883
Db       2236 CCTGTATTCAACCTCGATGACATAGAAGAGGCAGTTGAAGGTGCTGGTTTTGATGATGAA 2295

Qy       1884 AGGTCAATGCTGATCTCCCAGTTGAGCTTTGAGAAAACTTTTGGCTTATCTTCTGTCTTC 1943
Db       2296 AAGGCGTCTTTAATGTCGCAAATGAGCCTGGAGAAGCGATTTGGACAGTCTGCTGTTTTT 2355

Qy       1944 ATTGAGT[illegible]TACACTAATGGAGAATGGA[illegible]AGTACCGAGTCTGCCAACTCACCACCATTC 2003
Db       2356 GTTGCTTCTACCCTAATGGAAAATGGTGGTGTTCCTCCTTCAGCAACTCCAGAAAACTTT 2415

Qy       2004 ATCAAGGAAGCGATT[illegible]AAGTCATCGGCTGTGGCTATGAAGAGAAGACTGAATGGGGAAAA 2063
Db       2416 CTCAAA[illegible]AGGCTATCGATGTCATTAGTTGTGGTTATGAGGATAAGTCAGATTGGGGAATG 2475

Qy       2064 CAGATT[illegible]TGGATATATGGGT[illegible]AGTCACTGAGGATATCTTAAGTGGCTTCAAGATGCAC 2123
Db       2476 GAGATT[illegible]GGATGGATCTATGGTTCTGTGACAGAAGATATTCTGACTGGGTTCAAAATGCAT 2535

Qy       2124 TGCCGAGGATGGAGATCAATTTACTGCATGCCCGTAAGGCCTGCATTCAAAGGATCTGCA 2183
Db       2536 GCCCGTGGATGGCGATCCATTTACTGCATGCCTAAGCTTCCAGCTTTCAAGGGTTCTGCT 2595

Qy       2184 CCCATCAACCTGTCTGATAGATTGCACCAGGTCCTCCGATGGGCTCTTGGTTCTGTGGAA 2243
Db       2596 CCTATCAATCTTTCAGATCGTCTGAACCAAGTGCTGAGGTGGGCTTTAGGTTCAGTTGAG 2655
```



```
Qy       2244 ATTTTCTTTAGCAGACACTGTCCCCTCTGGTACGGGTTTGGAGGAGGCCGTCTTAAATGG 2303
Db       2656 ATTCTCTTCAGTCGGCATTGTCCTATATGGTATGG---TTACAATGGGAGGCTAAAATTT 2712

Qy       2304 CTCCAAAGGCTTGCGTATATAAACACCATTGTGTACCCATTTACATCCCTCCCTCTCATT 2363
Db       2713 CTTGAGAGGTTTGCGTATGTGAACACCACCATCTACCCTATCACCTCCATTCCTCTTCTC 2772

Qy       2364 GCCTATTGCACAATTCCTGCAGTTTGTCTGCTCACCGGAAAATTCATCATACCAACGCTC 2423
Db       2773 ATGTATTGTACATTGCTAGCCGTTTGTCTCTTCACCAACCAGTTTATTATTCCTCAGATT 2832

Qy       2424 TCAAACCTGGCAAGCATGCTGTTTCTTGGCCTCTTTATCTCCATCATTGTAACTGCGGTG 2483
Db       2833 AGTAACATTGCAAGTATATGGTTTCTGTCTCTCTTTCTCTCCATTTTCGCCACGGGTATA 2892

Qy       2484 CTTGAGCTAAGATGGAGCGGTGTCAGCATTGAAGATTTATGGCGTAATGAACAATTCTGG 2543
Db       2893 CTAGAAATGAGGTGGAGTGGCGTAGGCATAGACGAATGGTGGAGAAACGAGCAGTTTTGG 2952

Qy       2544 GTGATCGGAGGTGTTTCAGCCCATCTCTTTGCGGTCTTCCAGGGATTCTTAAAAATGTTG 2603
Db       2953 GTCATTGGTGGAGTATCCGCTCATTTATTCGCTGTGTTTCAAGGTATCCTCAAAGTCCTT 3012

Qy       2604 GCTGGCATCGATACGAACTTCACTGTCACAGCAAAAG---CAGCCGAAGATGCAGAATTT 2660
Db       3013 GCCGGTATTGACACAAACTTCACAGTTACCTCAAAAGCTTCAGATGAAGACGGAGACTTT 3072

Qy       2661 GGGGAGCTATATATGGTCAAGTGGACAACACTTTTGATTCCTCCAACCACACTTCTCATT 2720
Db       3073 GCTGAGCTCTACTTGTTCAAATGGACAACACTTCTGATTCCGCCAACGACGCTGCTCATT 3132

Qy       2721 ATCAATATGTCGGG---TTGTGCTGGATTCTCTGATGCACTCAACAAAGGATATGAAGCA 2777
Db       3133 GTAAACTTAGTGGGAGTTGTTGCAGGAGTCTCTTATGCTATCAACAGTGGATACCAATCA 3192

Qy       2778 TGGGGGCCTCTCTTTGGCAAGGTGTTCTTTGCTTTCTGGGTGATTCTTCATCTCTATCCA 2837
Db       3193 TGGGGACCACTCTTTGGTAAGTTGTTCTTTGCTTTCTGGGTGATTGTTCACTTGTACCCT 3252

Qy       2838 TTCCTTAAAGGTCTAATGGGTGGCAAAACCTAACACCAACCATTGTTGTTCTCTGGTCA 2897
Db       3253 TTCCTCAAGGGTTTGATGGGTCGACAGAACCGGACTCCTACCATTGTTGTGGTCTGGTCT 3312

Qy       2898 GTGCTGTTGGCCTCTGTCTTCTCTCTGTTTGGGTCAAGATCAATCCATTCGTTAACAAA 2957
Db       3313 GTTCTCTTGGCTTCTATCTTCTCGTTTTGTGGTTACGATTGATCCTTCACTAGCCGA 3372

Qy       2958 GTTGATAACACCTTGGTTGCGA ...
Db       3373 GTCACTGACCCGGACATTCTCGA ...
```

```
RESULT 15
BD022678
LOCUS       BD022678               3614 bp    DNA     linear   PAT 27 AUG 2002
DEFINITION  Manipulation of cellulose and/or beta-1,4-glucan.
ACCESSION   BD022678
VERSION     BD022678.1  GI:22563901
KEYWORDS    JP 2001510326-A/6.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 3614)
  AUTHORS   Ariori,A., Williamson,R.E., Bettsunar,A.S. and Penn,L.
  TITLE     Manipulation of cellulose and/or beta-1,4 glucan
  JOURNAL   Patent: JP 2001510326-A 6 31-JUL-2001;
            THE AUSTRALIAN NATIONAL UNIVERSITY, COMMONWEALTH SCIENTIFIC AND
            INDUSTRIAL RESEARCH ORGANIZATION
COMMENT     OS   Arabidopsis thaliana (thale cress)
            PN   JP 2001510326-A/6
            PD   31-JUL-2001
            PF   24-JUN-1997 JP 1998503665
            PR   27-JUN-1996 AU      PO 0699
            PI   ANTONIO ARIORI,RICHARD EDWARD WILLIAMSON,ANDREAS STEPHAN  PI
            BETTSUNAR,
            PI   LYONKAI PENN
            PC   C12N15/09,A01H5/00,C07K16/40,C12N1/15,C12N1/19,C12N1/21,C12N5/
            PC   10,C12N5/10,
            PC   C12N9/10,C12N15/00,C12N5/00,C12N5/00
            CC   Strandedness: Single;
            CC   Topology: Linear;
            FH   Key             Location/Qualifiers
            FT   CDS             217. .3411.
FEATURES             Location/Qualifiers
     source          1. .3614
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3702"
BASE COUNT      963 a    717 c    863 g   1071 t
ORIGIN
```

Query Match 35.1%; Score 1133.8; DB 6; Length 3614;
Best Local Similarity 68.1%; Pred. No. 1.6e-273;
Matches 1691; Conservative 0; Mismatches 762; Indels 30; Gaps 7;

```
Qy        519 CCGTCTGCAGAGGCTTCGGAGCCGCTTTCAATTGTTTATCCAATTCCACGCAACAAGCTC 578
Db        922 CTGAATGACGAAGCGAGGCAGCTTCTGTCAAGGAAAGTTTCAATTCCTTCATCACGGATC 981

Qy        579 ACACCATACAGAGCAGTGATCATTATGCGACTGGTCATTCTGGGCCTCTTCTTCCACTTC 638
Db        982 AATCCTTACAGAATGGTTATTATGCTGCGGCTTGTTATCCTTTGTCTCTTCTTGCATTAC 1041

Qy        639 AGAATAACAAATCCTGTCGATAGTGCCTTTGGCCTGTGGCTTACTTCTGTCATATGTGAG 698
Db       1042 CGTATAACAAACCCAGTGCCAAATGCCTTTGCTCTATGGCTGGTCTCTGTGATATGTGAG 1101

Qy        699 ATCTGGTTTGCATTTTCTTGGGTGTTGGATCAGTTCCCCAAGTGGAATCCTGTCAATAGA 758
Db       1102 ATCTGGTTTGCCTTATCCTGGATTTTGGATCAGTTTCCCAAGTGGTTTCCTGTGAACCGT 1161

Qy        759 GAAACGTATATCGAAAGGCTGTCGGCAAGGTATGAAAGAGAGGGTGAGCCTTCTCAGCTT 818
Db       1162 GAAACCTACCTCGACAGGCTTGCTTTAAGATATGATCGTGAAGGTGAGCCATCACAGTTA 1221

Qy        819 GCTGGTGTGGATTTTTTCGTGAGTACTGTTGATCCGCTGAAGGAACCGCCATTGATCACT 878
Db       1222 GCTGCTGTTGACATTTTCGTGAGTACTGTTGACCCCTTGAAGGAGCCACCCCTTGTGACA 1281

Qy        879 GCCAATACAGTCCTTTCCATCCTTGCTGTGGACTATCCCGTCGATAAAGTCTCCTGCTAC 938
Db       1282 GCCAACACAGTGCTCTCTATTCTGGCTGTTGACTACCCAGTTGACAAGGTGTCCTGTTAT 1341

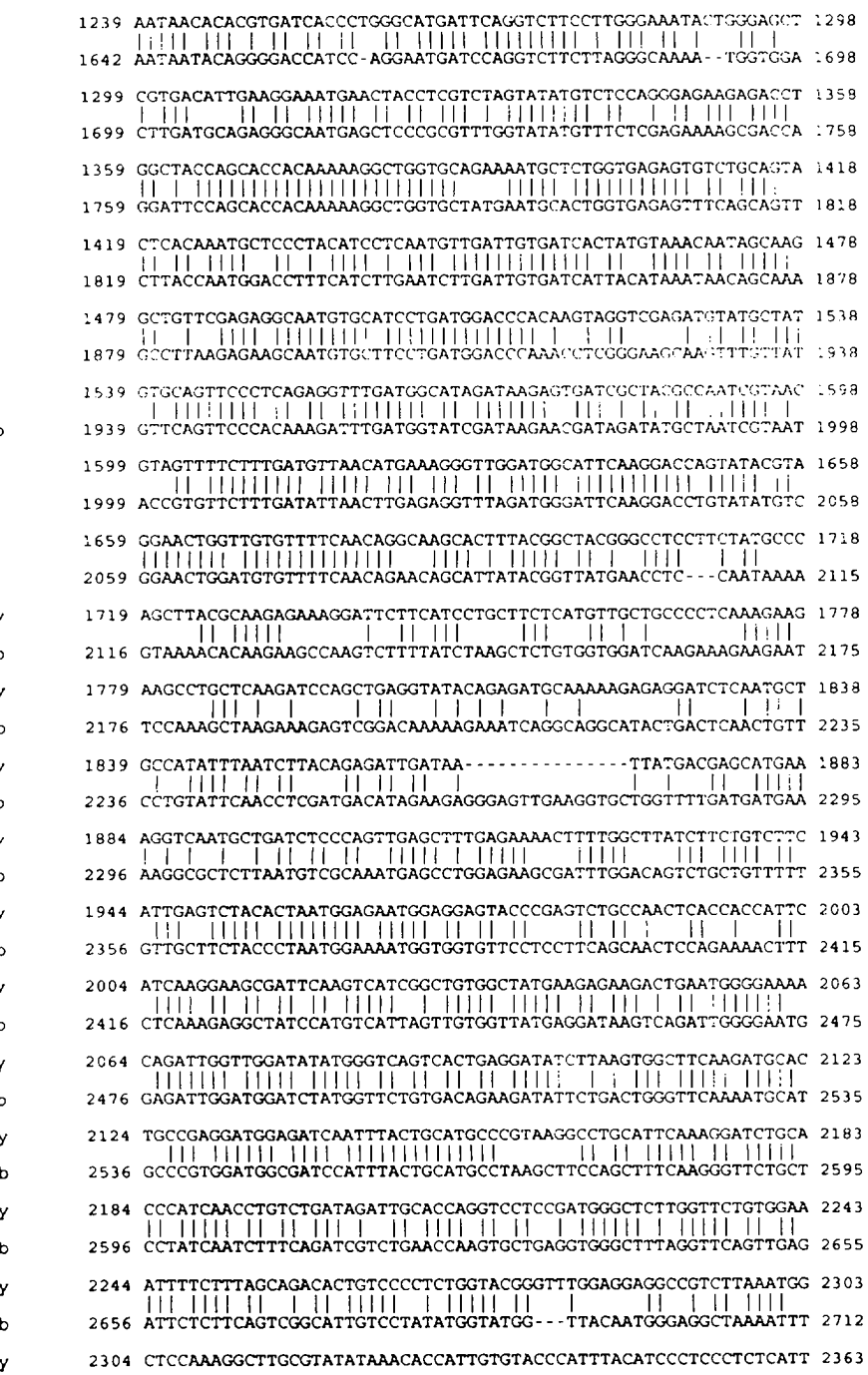
Qy        939 GTGTCTGATGATGGTGCAGCTATGCTTTCATTTGAATCTCTTGTAGAAACAGCTGAGTTT 998
Db       1342 GTTTCTGATGATGGTGCTGCTATGTTATCATTTGAATCACTTGCAGAAACATCAGAGTTT 1401

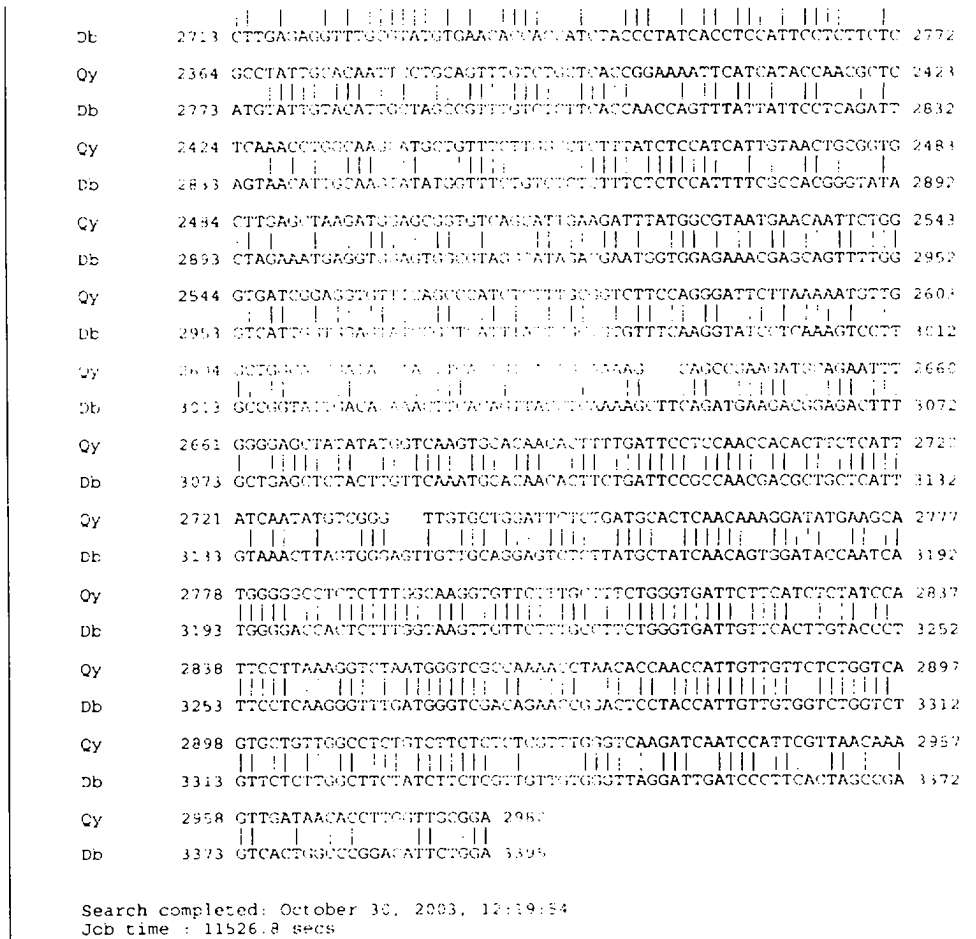
Qy        999 GCAAGGAAGTGGGTTCCGTTCTGCAAAAAATTCTCAATTGAACCAAGAGCACCGGAGTTT 1058
Db       1402 GCTCGTAAATGGGTACCATTTTGCAAGAAATATAGCATAGAGCCTCGTGCACCAGAATGG 1461

Qy       1059 TACTTCTCACAGAAAATTGATTACTTGAAAGACAAGGTTCAACCTTCTTTCGTGAAAGAA 1118
Db       1462 TACTTTGCTGCGAAAATAGATTACTTGAAGGATAAAGTTCAGACATCATTTGTCAAAGAT 1521

Qy       1119 CGTAGAGCAATGAAAAGGGATTATGAAGAGTACAAAGTCCGAGTTAATGCCCTGGTAGCA 1178
Db       1522 CGTAGAGCTATGAAGAGGGAATATGAGGAATTTAAAATCCGAATCAATGCACTTGTTTCC 1581

Qy       1179 AAGGCTCAGAAAACACCTGAAGAAGGATGGACTATGCAAGATGGAACACCTTGGCCTGGG 1238
Db       1582 AAAGCCCTAAAATGTCCTGAAGAAGGGTGGGTTATGCAAGATGGCACACCGTGGCCTGGA 1641
```





Search completed: October 30, 2003, 12:19:54
Job time : 11526.8 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 30, 2003, 05:34:29 ; Search time 785.73 Seconds
(without alignments)
11103.788 Million cell updates/sec

Title: US-09-980-043-1
Perfect score: 3232
Sequence: 1 gtcgacccacgcgtccgtct.....aaaaaaaaaaaaaaaaaaa 3232

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3232	100.0	3232	22	AAC65448	Populus tremuloide
2	1991.4	61.6	3328	19	AAV34432	Cotton cellulose s
3	1988.2	61.5	3207	19	AAV08372	Cellulose synthase
4	1848.6	57.2	3747	21	AAA67145	Eucalyptus grandis
5	1173	36.3	3851	21	AAA67114	Pinus radiata cell
6	1152.6	35.7	3517	21	AAZ58268	Soybean cellulose
7	1148.8	35.5	3753	21	AAZ99515	DNA encoding a mai
8	1148.8	35.5	3753	21	AAZ99530	DNA encoding a mai
9	1148.8	35.5	3780	21	AAZ99497	DNA encoding a mai
10	1137.8	35.2	3704	21	AAZ99533	DNA encoding a mai
11	1137.6	35.2	3725	21	AAZ99500	DNA encoding a mai
12	1137.6	35.2	3725	21	AAZ99506	DNA encoding a mai
13	1137.6	35.2	3725	21	AAZ99521	DNA encoding a mai
14	1135.4	35.1	3198	21	AAC49550	Arabidopsis thalia
15	1133.8	35.1	3614	19	AAV06567	Arabidopsis cellul
16	1128.6	34.9	3799	24	AAS16458	Corn cDNA encoding
17	1106.2	34.2	3936	21	AAZ58266	Corn cellulose syn
18	1106.2	34.2	3969	21	AAZ99503	DNA encoding a mai
19	1106.2	34.2	3969	21	AAZ99518	DNA encoding a mai
20	1104.8	34.2	3311	19	AAV08373	Cellulose synthase
21	1103.2	34.1	2830	24	AAS16455	Corn cDNA encoding
22	1102.8	34.1	3746	21	AAZ99512	DNA encoding a mai
23	1102.8	34.1	3746	21	AAZ99527	DNA encoding a mai
24	1102.8	34.1	3773	21	AAZ99494	DNA encoding a mai
25	1097.8	34.0	3603	19	AAV06565	Arabidopsis cellul
26	1097	33.9	3568	21	AAZ99491	DNA encoding a mai
27	1097	33.9	3776	21	AAZ58263	Corn cellulose syn
28	1093	33.8	3673	19	AAV06568	Arabidopsis cellul
29	1082	33.5	3786	21	AAZ58265	Corn cellulose syn
30	1082	33.5	3813	21	AAZ99509	DNA encoding a mai
31	1082	33.5	3813	21	AAZ99524	DNA encoding a mai
32	1073	33.2	2890	21	AAZ58270	Soybean cellulose
33	1051	32.5	3255	24	ABZ12754	Arabidopsis thalia
34	1048.6	32.4	3255	24	ABZ14725	Arabidopsis thalia
35	1043	32.3	3828	19	AAV06566	Arabidopsis cellul
36	1039	32.1	3444	22	AAC83798	Arabidopsis thalia
37	923.4	28.6	1407	21	AAA67086	Eucalyptus grandis
38	918.2	28.4	2306	19	AAV34433	Cotton cellulose s
39	853.2	26.4	2033	19	AAV08381	Cellulose synthase
40	775.4	24.0	2248	19	AAV06562	Arabidopsis cellul
41	764.4	23.7	2125	21	AAZ58269	Soybean cellulose
42	758.2	23.5	2055	21	AAA67111	Eucalyptus grandis
43	651.4	20.2	1860	21	AAA67081	Eucalyptus grandis
44	636.8	19.7	1734	21	AAZ58271	Wheat cellulose sy
45	550	17.0	950	21	AAC50201	Arabidopsis thalia

ALIGNMENTS

RESULT 1
AAC65448
ID AAC65448 standard; cDNA; 3232 BP.
XX
AC AAC65448;
XX
DT 07-MAR 2001 (first entry)
XX
DE Populus tremuloides CelA cDNA.
XX
KW Aspen; CelA; cellulose synthase; arabidopsis; wood;
KW mechanical stress response element; MSRE; cellulose biosynthesis; ss.
XX
OS Populus tremuloides.
XX
PN WO200071670-A2.
XX
PD 30-NOV-2000.
XX
PF 18-MAY-2000; 2000WO-US13637.
XX
PR 21-MAY-1999; 99US-0135280.
XX
PA (UNMT) UNIV MICHIGAN TECHNOLOGICAL.
XX
PI Chiang VL, Wu L, Chandrashekhar JP, Carraway DT;
XX
DR WPI; 2001-032025/04.
DR P-PSDB; AAB37893.
XX
PT Novel polynucleotide encoding cellulose synthase useful for enhancing

cellulose biosynthesis, modifying lignin biosynthesis, and for accelerating growth of plants and for upregulating cellulose synthase levels -

Claim 1; Fig 1; 65pp; English.

The present sequence is given in a specification relating to polynucleotides encoding cellulose synthase enzymes from Populus tremuloides and Arabidopsis. The cellulose synthases are designated PtCelA and AraxCelA respectively. The polynucleotides may be introduced into plants where they are useful for altering the growth of a plant. Vectors comprising a cellulose synthase promoter, or a fragment of the promoter containing one or more mechanical stress response elements (MSRE), operably linked to a cellulose synthase gene are useful for causing stress-induced gene expression in a plant cell. A cDNA encoding a protein that binds to a positive MSRE may be delivered into the plant to increase expression of cellulose synthase. A cDNA in an antisense orientation may be delivered to reduce expression of cellulose synthase. The polynucleotides enable the development of transgenic tree varieties having increased cellulose content, decreased lignin content and therefore improved wood fibre characteristics. The cellulose synthase promoters are useful for synthesising greater quantities of high-crystalline cellulose in plant and preferably in trees, which permits an increase in proportion of cellulose in transgenic plants, greater strength of juvenile wood and fiber and acceleration of overall growth rate.

Sequence 3232 BP; 895 A; 668 C; 772 G; 897 T; 0 other;

Query Match 100.0%; Score 3232; DB 22; Length 3232;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3232; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    1 GTCGACCCACGCGTCCGTCTTGAAAGAATATGAAGTTGTAAAGAGCTGGTAAAGTGGTAA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 GTCGACCCACGCGTCCGTCTTGAAAGAATATGAAGTTGTAAAGAGCTGGTAAAGTGGTAA 60

Qy   61 TAAGCAAGATGATGGAATCTGGGGCTCCTATATGCCATACCTGTGGTGAACAGGTGGGGC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 TAAGCAAGATGATGGAATCTGGGGCTCCTATATGCCATACCTGTGGTGAACAGGTGGGGC 120

Qy  121 ATGATGCAAATGGGGAGCTATTTGTGGCTTGCCATGAGTGTAGCTATCCCATGTGCAAGT 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 ATGATGCAAATGGGGAGCTATTTGTGGCTTGCCATGAGTGTAGCTATCCCATGTGCAAGT 180

Qy  181 CTTGTTTCGAGTTTGAAATCAATGAGGGCCGGAAAGTTTGCTTGCGGTGTGGCTCGCCAT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 CTTGTTTCGAGTTTGAAATCAATGAGGGCCGGAAAGTTTGCTTGCGGTGTGGCTCGCCAT 240

Qy  241 ATGATGAGAACTTGCTGGATGATGTAGAAAAGAAGGGGTCTGGCAATCAATCCACAATGG 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 ATGATGAGAACTTGCTGGATGATGTAGAAAAGAAGGGGTCTGGCAATCAATCCACAATGG 300

Qy  301 CATCTCACCTCAACGATTCTCAGGATGTCGGAATCCATGCTAGACATATCAGTAGTGTGT 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 CATCTCACCTCAACGATTCTCAGGATGTCGGAATCCATGCTAGACATATCAGTAGTGTGT 360

Qy  361 CCACTGTGGATAGTGAAATGAATGATGAATATGGGAATCCAATTTGGAAGAATCGGGTGA 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 CCACTGTGGATAGTGAAATGAATGATGAATATGGGAATCCAATTTGGAAGAATCGGGTGA 420

Qy  421 AGAGCTGTAAGGATAAAGAGAACAAGAAGAAAAAAGAGAAGTCCTAAGGCTGAAACTGAAC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 AGAGCTGTAAGGATAAAGAGAACAAGAAGAAAAAAGAGAAGTCCTAAGGCTGAAACTGAAC 480

Qy  481 CAGCTCAAGTTCCTACAGAACAGCAGATGGAAGAGAAACCGTCTGCAGAGGCTTCGGAGC 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481 CAGCTCAAGTTCCTACAGAACAGCAGATGGAAGAGAAACCGTCTGCAGAGGCTTCGGAGC 540

Qy  541 CGCTTTCAATTGTTTATCCAATTCCACGCAACAAGCTCACACCATACAGAGCAGTGATCA 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541 CGCTTTCAATTGTTTATCCAATTCCACGCAACAAGCTCACACCATACAGAGCAGTGATCA 600

Qy  601 TTATGCGACTGGTCATTCTGGGCCTCTTCTTCCACTTCAGAATAACAAATCCTGTCGATA 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601 TTATGCGACTGGTCATTCTGGGCCTCTTCTTCCACTTCAGAATAACAAATCCTGTCGATA 660

Qy  661 GTGCCTTTGGCCTGTGGCTTACTTCTGTCATATGTGAGATCTGGTTTGCATTTTCTTGGG 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  661 GTGCCTTTGGCCTGTGGCTTACTTCTGTCATATGTGAGATCTGGTTTGCATTTTCTTGGG 720

Qy  721 TGTTGGATCAGTTCCTCAAGTGGAATCCTGTCAATAGAGAAACGTATATCGAAAGGCTGT 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  721 TGTTGGATCAGTTCCTCAAGTGGAATCCTGTCAATAGAGAAACGTATATCGAAAGGCTGT 780

Qy  781 CGGCAAGGTATGAAAAGAGAGGGTGAGCCTTCTCAGCTTGCTGGTGTGGATTTTTTCGTGA 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  781 CGGCAAGGTATGAAAAGAGAGGGTGAGCCTTCTCAGCTTGCTGGTGTGGATTTTTTCGTGA 840

Qy  841 GTACTGTTGATCCGCTGAAGGAACCGCCATTGATCACTGCCAATACAGTCCTTTCCATCC 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  841 GTACTGTTGATCCGCTGAAGGAACCGCCATTGATCACTGCCAATACAGTCCTTTCCATCC 900

Qy  901 TTGCTGCGGACTATCCGGTCGATAAAGTCGCCGTACGTGTCTGATGATGGTGCAGCTA 960
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  901 TTGCTGCGGACTATCCGGTCGATAAAGTCGCCGTACGTGTCTGATGATGGTGCAGCTA 960

Qy  961 TGCTTTCATTTGAATCTCTTGTAGAAACAGCTGAGTTTGCAAGGAAGTGGGTTCCGTTCT 1020
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  961 TGCTTTCATTTGAATCTCTTGTAGAAACAGCTGAGTTTGCAAGGAAGTGGGTTCCGTTCT 1020

Qy 1021 GCAAAAAATTCTCAATTGAACCAAGAGCACCGGAGTTTTACTTCTCACAGAAAATTGATT 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1021 GCAAAAAATTCTCAATTGAACCAAGAGCACCGGAGTTTTACTTCTCACAGAAAATTGATT 1080

Qy 1081 ACTTGAAAGACAAGGTTCAACCTTCTTTCGTGAAAGAACGTAGAGCAATGAAAAGGGATT 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1081 ACTTGAAAGACAAGGTTCAACCTTCTTTCGTGAAAGAACGTAGAGCAATGAAAAGGGATT 1140

Qy 1141 ATGAAGAGTACAAAGTCCGAGTTAATGCCGTGGTAGCAAAGGCTCAGAAAACACCTGAAG 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1141 ATGAAGAGTACAAAGTCCGAGTTAATGCCGTGGTAGCAAAGGCTCAGAAAACACCTGAAG 1200

Qy 1201 AAGGATGGACTATGCAAGATGGAACACCTTGGCCTGGGAATAACACACGTGATCACCCTG 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1201 AAGGATGGACTATGCAAGATGGAACACCTTGGCCTGGGAATAACACACGTGATCACCCTG 1260

Qy 1261 GGCATGATTCAGGTCTTCCTTGGGAAATACTGGGAGCTCGTGACATTGAAGGAAATGAAC 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1261 GGCATGATTCAGGTCTTCCTTGGGAAATACTGGGAGCTCGTGACATTGAAGGAAATGAAC 1320

Qy 1321 TACCTCGTCTAGTATATGTCTCCAGGGAGAAGAGACCTGGCTACCAGCACCACAAAAAGG 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1321 TACCTCGTCTAGTATATGTCTCCAGGGAGAAGAGACCTGGCTACCAGCACCACAAAAAGG 1380

Qy 1381 CTGGTGCAGAAAATGCTCTGGTGAGAGTGTCTGCAGTACTCACAAATGCTCCCTACATCC 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1381 CTGGTGCAGAAAATGCTCTGGTGAGAGTGTCTGCAGTACTCACAAATGCTCCCTACATCC 1440

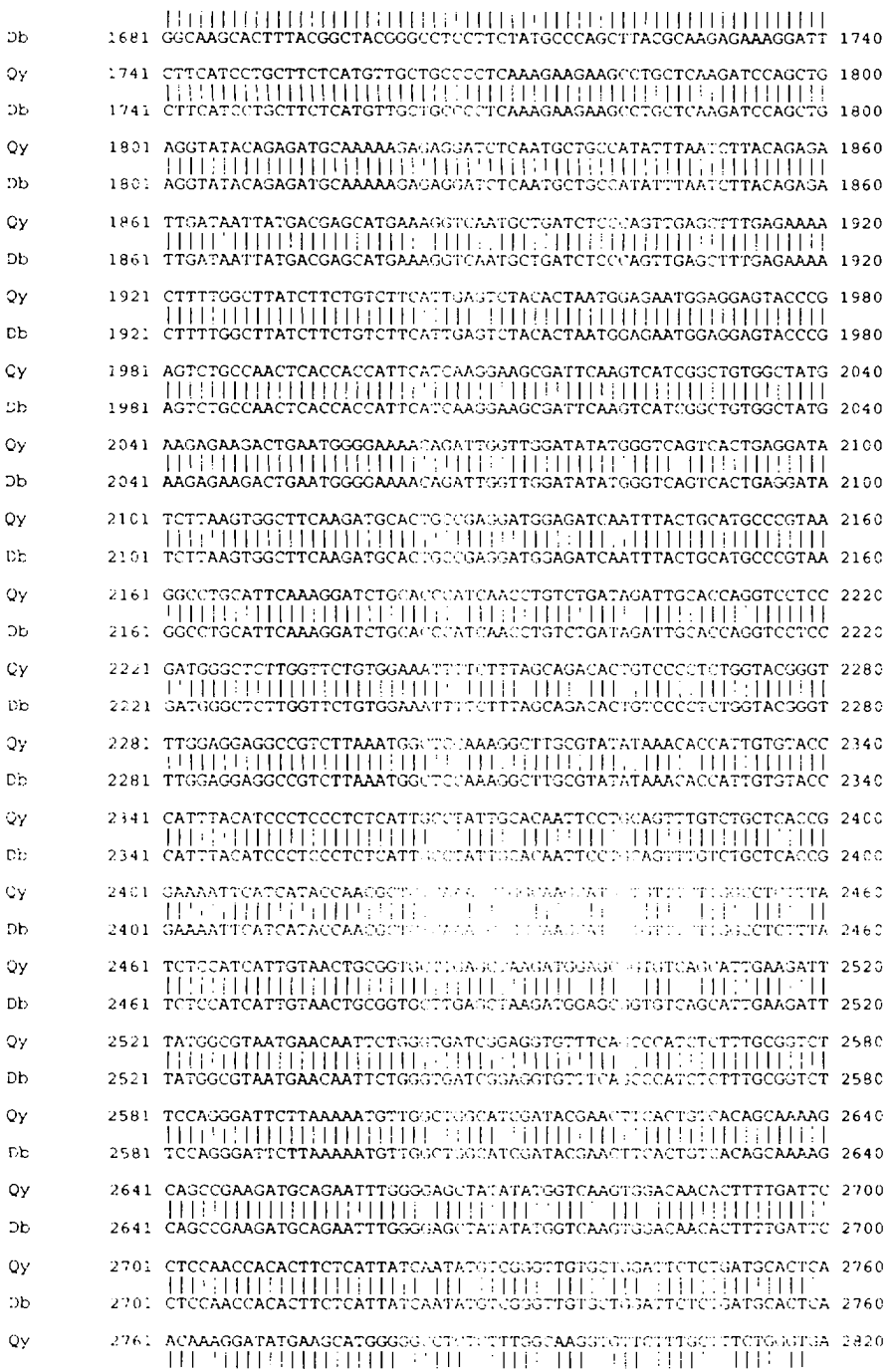
Qy 1441 TCAATGTTGATTGTGATCACTATGTAAACAATAGCAAGGCTGTTCGAGAGGCAATGTGCA 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1441 TCAATGTTGATTGTGATCACTATGTAAACAATAGCAAGGCTGTTCGAGAGGCAATGTGCA 1500

Qy 1501 TCCTGATGGACCCACAAGTAGGTCGAGATGTATGCTATGTGCAGTTCCCTCAGAGGTTTG 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1501 TCCTGATGGACCCACAAGTAGGTCGAGATGTATGCTATGTGCAGTTCCCTCAGAGGTTTG 1560

Qy 1561 ATGGCATAGATAAGAGTGATCGCTACGCCAATCGTAACGTAGTTTTCTTTGATGTTAACA 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1561 ATGGCATAGATAAGAGTGATCGCTACGCCAATCGTAACGTAGTTTTCTTTGATGTTAACA 1620

Qy 1621 TGAAAGGGTTGGATGGCATTCAAGGACCAGTATACGTAGGAACTGGTTGTGTTTTCAACA 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1621 TGAAAGGGTTGGATGGCATTCAAGGACCAGTATACGTAGGAACTGGTTGTGTTTTCAACA 1680

Qy 1681 GGCAAGCACTTTACGGCTACGGGCCTCCTTCTATGCCCAGCTTACGCAAGAGAAAGGATT 1740
```



```
Db       1681 GGCAAGCACTTTACGGCTACGGGCCTCCTTCTATGCCCAGCTTACGCAAGAGAAAGGATT 1740

Qy       1741 CTTCATCCTGCTTCTCATGTTGCTGCCCCTCAAAGAAGAAGCCTGCTCAAGATCCAGCTG 1800

Db       1741 CTTCATCCTGCTTCTCATGTTGCTGCCCCTCAAAGAAGAAGCCTGCTCAAGATCCAGCTG 1800

Qy       1801 AGGTATACAGAGATGCAAAAAGAGAGGATCTCAATGCTGCCATATTTAATCTTACAGAGA 1860

Db       1801 AGGTATACAGAGATGCAAAAAGAGAGGATCTCAATGCTGCCATATTTAATCTTACAGAGA 1860

Qy       1861 TTGATAATTATGACGAGCATGAAAGGTCAATGCTGATCTCCCAGTTGAGCTTTGAGAAAA 1920

Db       1861 TTGATAATTATGACGAGCATGAAAGGTCAATGCTGATCTCCCAGTTGAGCTTTGAGAAAA 1920

Qy       1921 CTTTTGGCTTATCTTCTGTCTTCATTGAGTCTACACTAATGGAGAATGGAGGAGTACCCG 1980

Db       1921 CTTTTGGCTTATCTTCTGTCTTCATTGAGTCTACACTAATGGAGAATGGAGGAGTACCCG 1980

Qy       1981 AGTCTGCCAACTCACCACCATTCATCAAGGAAGCGATTCAAGTCATCGGCTGTGGCTATG 2040

Db       1981 AGTCTGCCAACTCACCACCATTCATCAAGGAAGCGATTCAAGTCATCGGCTGTGGCTATG 2040

Qy       2041 AAGAGAAGACTGAATGGGGAAAACAGATTGGTTGGATATATGGGTCAGTCACTGAGGATA 2100

Db       2041 AAGAGAAGACTGAATGGGGAAAACAGATTGGTTGGATATATGGGTCAGTCACTGAGGATA 2100

Qy       2101 TCTTAAGTGGCTTCAAGATGCACTGCCGAGGATGGAGATCAATTTACTGCATGCCCGTAA 2160

Db       2101 TCTTAAGTGGCTTCAAGATGCACTGCCGAGGATGGAGATCAATTTACTGCATGCCCGTAA 2160

Qy       2161 GGCCTGCATTCAAAGGATCTGCACCCATCAACCTGTCTGATAGATTGCACCAGGTCCTCC 2220

Db       2161 GGCCTGCATTCAAAGGATCTGCACCCATCAACCTGTCTGATAGATTGCACCAGGTCCTCC 2220

Qy       2221 GATGGGCTCTTGGTTCTGTGGAAATTTTCTTTAGCAGACACTGTCCCCTCTGGTACGGGT 2280

Db       2221 GATGGGCTCTTGGTTCTGTGGAAATTTTCTTTAGCAGACACTGTCCCCTCTGGTACGGGT 2280

Qy       2281 TTGGAGGAGGCCGTCTTAAATGGCTCCAAAGGCTTGCGTATATAAACACCATTGTGTACC 2340

Db       2281 TTGGAGGAGGCCGTCTTAAATGGCTCCAAAGGCTTGCGTATATAAACACCATTGTGTACC 2340

Qy       2341 CATTTACATCCCTCCCTCTCATTGCCTATTGCACAATTCCTGCAGTTTGTCTGCTCACCG 2400

Db       2341 CATTTACATCCCTCCCTCTCATTGCCTATTGCACAATTCCTGCAGTTTGTCTGCTCACCG 2400

Qy       2401 GAAAATTCATCATACCAAGGCTCTAAAGTTAAAAAGCATTTTTCTTTGCCCTTTTTA 2460

Db       2401 GAAAATTCATCATACCAAGGCTCTAAAGTTAAAAAGCATTTTTCTTTGCCCTTTTTA 2460

Qy       2461 TCTCCATCATTGTAACTGCGGTGCTTGAGCTAAGATGGAGCTGTGTCAGCATTGAAGATT 2520

Db       2461 TCTCCATCATTGTAACTGCGGTGCTTGAGCTAAGATGGAGCTGTGTCAGCATTGAAGATT 2520

Qy       2521 TATGGCGTAATGAACAATTCTGGGTGATCGGAGGTGTTTCAGCCCATCTCTTTGCGGTCT 2580

Db       2521 TATGGCGTAATGAACAATTCTGGGTGATCGGAGGTGTTTCAGCCCATCTCTTTGCGGTCT 2580

Qy       2581 TCCAGGGATTCTTAAAAATGTTGGCTGGCATCGATACGAACTTCACTGTCACAGCAAAAG 2640

Db       2581 TCCAGGGATTCTTAAAAATGTTGGCTGGCATCGATACGAACTTCACTGTCACAGCAAAAG 2640

Qy       2641 CAGCCGAAGATGCAGAATTTGGGGAGCTATATATGGTCAAGTGGACAACACTTTTGATTC 2700

Db       2641 CAGCCGAAGATGCAGAATTTGGGGAGCTATATATGGTCAAGTGGACAACACTTTTGATTC 2700

Qy       2701 CTCCAACCACACTTCTCATTATCAATATGTCGGGTTGTGCTGGATTCTCTGATGCACTCA 2760

Db       2701 CTCCAACCACACTTCTCATTATCAATATGTCGGGTTGTGCTGGATTCTCTGATGCACTCA 2760

Qy       2761 ACAAAGGATATGAAGCATGGGGGCCTCTCTTTGGCAAGGTGTTCTTTGCTTTCTGGGTGA 2820

Db       2761 ACAAAGGATATGAAGCATGGGGGCCTCTCTTTGGCAAGGTGTTCTTTGCTTTCTGGGTGA 2820

Qy       2821 TTCTTCATCTCTATCCATTCCTTAAAGGTCTAATGGGTCGCCAAAACCTAACACCAACCA 2880

Db       2821 TTCTTCATCTCTATCCATTCCTTAAAGGTCTAATGGGTCGCCAAAACCTAACACCAACCA 2880

Qy       2881 TTGTTGTTCTCTGGTCAGTGCTGTTGGCCTCTGTCTTCTCTCTCGTTTGGGTCAAGATCA 2940

Db       2881 TTGTTGTTCTCTGGTCAGTGCTGTTGGCCTCTGTCTTCTCTCTCGTTTGGGTCAAGATCA 2940

Qy       2941 ATCCATTCGTTAACAAAGTTGATAACACCTTGGTTGCGGAGACCTGCATTTCCATTGATT 3000

Db       2941 ATCCATTCGTTAACAAAGTTGATAACACCTTGGTTGCGGAGACCTGCATTTCCATTGATT 3000

Qy       3001 GCTGAGCTACCTCCAATAAGTCTCTCCCAGTATTTTGGGGTTACAAAACCTTTGGGAATT 3060

Db       3001 GCTGAGCTACCTCCAATAAGTCTCTCCCAGTATTTTGGGGTTACAAAACCTTTGGGAATT 3060

Qy       3061 GGAATATGATCCTCGTTGTAGTTTCCCTCAAGAAAGCACATATCGCTGTCAGTATTTAAA 3120

Db       3061 GGAATATGATCCTCGTTGTAGTTTCCCTCAAGAAAGCACATATCGCTGTCAGTATTTAAA 3120

Qy       3121 TGAACTGCAAGATGATTGTTCTCTATGAAGTTTTGAACAGTTTGAAATGATATTATGTTA 3180

Db       3121 TGAACTGCAAGATGATTGTTCTCTATGAAGTTTTGAACAGTTTGAAATGATATTATGTTA 3180

Qy       3181 AAATACAGGTTTTGATTGTGTTGAAAAAAAAAAAGAAAAAAAAAAAAAAAAA 3232

Db       3181 AAATACAGGTTTTGATTGTGTTGAAAAAAAAAAAGAAAAAAAAAAAAAAAAA 3232
```

RESULT 2
AAV34432
ID AAV34432 standard; cDNA; 3328 BP.
XX
AC AAV34432;
XX
DT 28-SEP-1998 (first entry)
XX
DE Cotton cellulose synthase CelA1 cDNA.
XX
KW Cellulose synthase; CelA1 gene; cotton; forest tree; plant fibre;
KW wood; transgenic plant; ds.
XX
CS Gossypium hirsutum Acala SJ-2.
XX
FH Key   Location/Qualifiers
FT CDS   179..3103
FT   /*tag= a
XX
PN WO9818949-A2.
XX
PD 07-MAY-1998.
XX
PF 29-OCT-1997; 97WO US19529.
XX
PR 29-OCT-1996; 96US-0029987.
XX
PA (CALJ) CALGENE INC.
XX
PI Delmer DP, Pear JR, Stalker DM;
XX
DR WPI; 1998-272238/24.
DR P-PSDB; AAW60039.
XX
PT New isolated plant cellulose synthesis enzyme DNA - is used to
PT develop products for producing modified plant fibre products,
PT particularly for modifying cotton fibre and wood quality
XX
PS Claim 5; Fig 6A-E; 49pp; English.
XX
CC This full-length cDNA sequence corresponds to the cotton CelA1 gene
CC transcript and encodes a 973-amino acid cellulose synthase (see

AAW60039). 2 cDNA clones that exhibited small blocks of amino acid homology to proteins encoded by bacterial CelA genes were isolated from an unamplified cDNA library prepared from mRNA of fibres of cotton Acala SJ-2 harvested at 21 dpa (time of maximal secondary wall cellulose synthesis). These 2 clones were partially homologous at the nucleotide and amino acid levels and were designated CelA1 and CelA2 (see AAV34433), respectively. The 2 cotton CelA genes are both expressed in developing fibre at approximately 500 times their level of expression in other cotton tissues, and constitute approximately 1-2% of the 24 dpa fibre mRNA. The CelA1 gene promoter region (see AAV34434) is also claimed. The isolated cDNA clones are used in claimed methods for modifying fibre phenotype in a cotton plant and for modifying wood quality phenotype in a forest tree species. Lignin can be reduced by over-expressing the cellulose synthase product and diverting production into cellulose.

Sequence 3328 BP; 914 A; 690 C; 763 G; 961 T; 0 other;

Query Match 61.6%; Score 1991.4; DB 19; Length 3328;
Best Local Similarity 80.8%; Pred. No. 0;
Matches 2400; Conservative 0; Mismatches 551; Indels 18; Gaps 6;

```
          40 AAAGAGCTGGTAAAGTGGTAATAAGCAAGATGATGGAATCTGGGGCTCCTATATGCCATA 99
             || |||   |  ||||  |  |    |  ||||||||||||||||| |||  |  |||. |
         150 AAGGAGAAAGATAAGTACTTTTTTTGAGAATGATGGAATCTGGGGTTCCTGTTTGCCACA 209

         100 CCTGTGGTGAACAGGTGGGGCATGATGCAAATGGGGAGCTATTTGTGGCTTGCCATGAGT 159
             | ||||||||||| || |||    |||  ||||| || :  |||||||||||||||||| |
         210 CTTGTGGTGAACATGTTGGGTTGAATGTTAATGGTGAACCTTTTGTGGCTTGCCATGAAT 269

         160 GTAGCTATCCCATGTGCAAGTCTTGTTTCGAGTTTGAAATCAATGAGGGCCGGAAAGTTT 219
             |||  |  || || || |||  |||||| |||| |||  |  || |  || |||| ||
         270 GTAATTTCCCTATTTGTAAGAGTTGTTTTGAGTATGATCTTAAGGAAGGACGAAAAGCTT 329

         220 GCTTGCGGTGTGGCTCGCCATATGATGAGAACTTGCTGGATGATGTAGAAAAGAAGGGGT 279
             |||||||  |||||     |||||||||||  ||| || |||| |||||  ||    ||||
         330 GCTTGCGTTGTGGTAGTCCATATGATGAAAACCTGTTGGACGATGTCGA---GAAGGCCA 386

         280 CTGGCAATCAATCCACAATGGCATCTCACCTCAACGATTCTCAGGATGTCGGAATCCATG 339
             | ||| |||||||  ||||||||  | ||  | ||| | ||||||||||| |||||  ||||
         387 CCGGCGATCAATCGACAATGGCTGCACATTTGAACAAGTCTCAGGATGTTGGAATTCATG 446

         340 CTAGACATATCAGTAGTGTGTCCACTGTGGATAGTGAAATGAATGATGAATATGGGAATC 399
             | |||||||||||  ||||||||  ||  ||||||||||||  ||| ||  |||||||||
         447 CAAGACATATCAGCAGTGTGTCTACATTGGATAGTGAAATGGCTGAAGACAATGGGAATT 506

         400 CAATTTGGAAGAATCGGGTGAAGAGCTGTAAGGATAAAGAGAACAAGAAGAAAAAGAGAA 459
             | ||||||||||||  |||||  | || || || || ||  ||||||||||||| |||
         507 CGATTTGGAAGAACAGGGTGGAAAGTTGGAAAGAAAAGAAGAACAAGAAGAAGAAGCCTG 566

         460 GTCCTAAGGCTGAAACTGAACCAGCTCAAGTTCCTACAGAACAGCAGATGGAAGAGAAAC 519
               | |      |    |  ||| || | ||  | || || || ||||||| ||||
         567 CAACAACTAAGGTTGAAAGAGAGGCTGAAATCCCACCTGAGCAACAAATGGAAGATAAAC 626

         520 CGTCTGCAGAGGCTTCGGAGCCGCTTTCAATTGTTTATCCAATTCCACGCAACAAGCTCA 579
             || |  | || |||||  |||| || || | |  |  ||||| ||   || ||   ||
         627 CGGCACCGGATGCTTCCCAGCCCCTCTCGACTATAATTGCAATCCCGAAAAGCAGACTTG 686

         580 CACCATACAGAGCAGTGATCATTATGCGACTGGTCATTCTGGGCCTCTTCTTCCACTTCA 639
             |||||||| || | ||||||||||||||| || ||||||| || || ||||||||  |
         687 CACCATACCGAACCGTGATCATTATGCGATTGATCATTCTTGGTCTTTTCTTCCATTATC 746

         640 GAATAACAAATCCTGTCGATAGTGCCTTTGGCCTGTGGCTTACTTCTGTCATATGTGAGA 699
             || ||||||| || || || ||||| |||||  |||||||  |||||  |||||||||||| |
         747 GAGTAACAAACCCCGTTGACAGTGCTTTTGGACTGTGGCTCACTTCAGTCATATGTGAAA 806

         700 TCTGGTTTGCATTTTCTTGGGTGTTGGATCAGTTCCCCAAGTGGAATCCTGTCAATAGAG 759
             ||||||||||||||| |||||||||||||||||||| ||||||  |||||||  || || |
         807 TCTGGTTTGCATTTTCCTGGGTGTTGGATCAGTTCCCTAAGTGGTATCCTGTTAACAGGG 866

         760 AAACGTATATCGAAAGGCTGTCGGCAAGGTATGAAAGAGAGGGTGAGCCTTCTCAGCTTG 819
             |||| || || || || || || ||||| |||||||||||  |||||  |||  | | ||||
         867 AAACATACATTGACAGACTATCTGCAAGATATGAAAGAGAAGGTGAACCTGATGAACTTG 926
```

```
Qy       820 CTGGTGTGGATTTTTCTGTGAGTATCTGCTTATTGGCTGAAGGAACCGCCATTGATCACTG 879
             ||| |  || ||| |||||| ||   |||     |||| || || |||||||| ||||
Db       927 CTGCAGTTGACTTCTTCGTGAGTATATGGGATTCATTGAAAGAGCCTCCATTGATTACTG 986

Qy       880 CCAATACAGTCCTTTCCATCCTTGCTGTGGAACTATCCCGTCGATAAAGTCTCCTGCTACG 939
             ||||||  || ||||||||||||  |  |||  || || ||||| |||||  ||  ''
Db       987 CCAATACTGTGCTTTCCATCCTTGCCTGGAACTACCCGGTGGATAAGGTCTCTTGTTATA 1046

Qy       940 TGTCTGATGATGGTGCAGCTATGCTTTCAGTTGAATCTCTTGTAGAAACAGCTGAGTTTG 999
             | |||||||||||  | |  |||  ||  ||||||||| ||||||||||| | ||||
Db      1047 TATCTGATGATGGTTGGGCATGCTGACATTTGAATCTCTAGTAGAAACAGCCGACTTTG 1106

Qy      1000 CAAGGAAGTGGGTTCTGTTCTGCAAAAAATTTTCAATTGAACCAAGAGCACCGGAGTTTT 1059
             '||, |||.|||. | |.|||||.|||  || |||||||  | .|||| |||||,
Db      1107 CAAGAAAGTGGGTTCTATTCTGCAAAAAATTTTTCATTGAACCCCGGGCACCTGAGTTTT 1166

Qy      1060 ACTTCTATAAAAAATGATTATTTTAAAAAAAGGTTCAACCTTCTTTCGGGAAAGAAG 1119
             .|||   |  ||  |||  ||  ||  ||  |||| |  |||| |  |||| |
Db      1167 ACTTCTATAAAAAATGATTATTTTAAAAAAAGGTTCAACCTTCTTTCGGGAAAGAAG 1226

Qy      1120 GTATAT AATAAAATTTATTATTAAAAA AAAGTTCGAGTTAATGTGTTTAGTAGCAA 1179
             ||| ||.  ||| .| .||.|.||  ||  |||  || |||  |  |||||
Db      1227 GTAGAGTTATGAAAAGAGATTATTAAGAGTAAAAATTCGAATCAATGCTTTAGTTGCAA 1286

Qy      1180 AGGCTCAGAAAAACATCTGAAGAACGATGGACTATGCAAGATGGAACACCTTGGGCTGGGA 1239
             |||||||  |||  |||  |||||  |||  |||||||||||  |.|||'  ||  '
Db      1287 AGGCTTAGAAAAACATTTGATGAATGATTGATAATGCAAGATGGAACTTCTTGGCCAGGAA 1346

Qy      1240 ATAACACACGTGATGACCCTGGGCATGATTCAAGTCTTCCTTGGGAAATACTGGGAGCTC 1299
             ||||||  ||||'|||'|||  ''|||  |||  | |||'|||  !  || |.|.
Db      1347 ATAACCTGCGTGATCACCCT GGCATGATTCAAGTTTTCCTTGG- ATATAGTGGTGCTC 1403

Qy      1300 GTGACATTGAAGGAAATGAACTTCCTCTGATTGTTATGTCTCCAGGGAGAAGAGACCTG 1359
             |.||||  ||||||||||||  |||.  | |   | |||.| || ||||||.||||||
Db      1404 GTGACATTGAAGGAAATGAACTTCCTCTGATTGTTACGTCTCTAGAGAGAAGAGACCTG 1463

Qy      1360 GCTACCAGCACCACAAAAAGGCTGGTGCTGAAAATGCTCTGGTGAGAGTGTCTGCAGTAC 1419
             |||.||| ,||||.|||||||||'|||  ||||||  |||| || '|||||||| |
Db      1464 GCTACCAACACCACAAAAAGGCTGGTGCTGAAAATGCTTTGGTTAGGGTGTCTGCAGTTC 1523

Qy      1420 TCACAAATGCTCCCTACATCCTCAATGTTGATTGTGATCACTATGTAAACAATAGCAAGG 1479
             | ||||'||||||  |||'|||.  |||.||||  |||||||. ||||.||||||||
Db      1524 TTACAAATGCTCCCTTCATCCTTAATGTTGATGTGACCACTATGTTAACAATAGCAAGG 1583

Qy      1480 CTGTTCGAGAGGCAATGTGCATGCTGATGGACCCACAAGTAGGTCGAGATGTATGCTATG 1539
             | ||| | ||||||||||||| '| |'|||||||||||| ||||||||||||||||||
Db      1584 CAGTTAGGGAGGCAATGTGCTTCTTGATGGACCCACAAGTTGGTCGAGATGTATGCTATG 1643

Qy      1540 TGCAGTTCGCCTCAGAGGTTTGATGGCATAGATAAGAGTGATCGCTACGCCAATCGTAACG 1599
             |||||||  |||| || '|||  |||| |||  |||||||| || ||||| | |.
Db      1644 TGCAGTTTCCTCAAAGATTTGATGGCATAGATAGGAGTGATCGATATGCCAATAGGAACA 1703

Qy      1600 TAGTTTTCTTTGATGTTAACATGAAAGGTTCGATGGCATTCAAGGACCAGTATACGTAG 1659
             ||  |||' |||  ||| ||||  ||| ||  |||| |||||  |||  || |. |
Db      1704 CAGTTTTCTTTGATGTTAACATGAAAGGTTCTTGATGGAATCCAAGGGCCAGTTTATGTGG 1763

Qy      1660 GAACTGGTTGTGTTTTCAACAGGCAAGCACTTTACGGCTACGGGCCTCCTTCTATGCCCA 1719
             .||| .|||| |||:||| ||| ;|||'|||| |||||  || || ||||| |||| |
Db      1764 GAACAGGTTGTGTTTTCAATAGGCAAGCACTTTATGGCTATGGTCCACCTTCAATGCCAA 1823

Qy      1720 GCTTACGCAAGAGAAAGGATTCTTCATCCTGTTTCTCATGTTGCTGCCCCTCAAAGAAGA 1779
             .|| ' |||| |      ;|  : ||  | |;| ||||||||||     |'||||
Db      1824 GTTTTCCCAAGTCA------TCCTCCTCATCTTGCTCGTGTTGCTGCCCGGGCAAGAAGG 1877

Qy      1780 AGCCTGCTCAAGATCCAGCTGAGGTATACAGAGATGCAAAAAGAGAGGATCTCAATGCTG 1839
             | |  ||  |||||||| | ||| | | ||  ||||||| |||  |  |||||
Db      1878 AAC---CTAAAGATCCATCAGAGCTTTATAGGGATGCAAAACGGGAAGAACTTGATGCTG 1934

Qy      1840 CCATATTTAATCTTACAGAGATTGATAATTATGACGAGCATGAAAGGTCAATGCTGATCT 1899
             |||| ||||| |||  || |||||  |||'||| ||| |||||||  ||||| |||||
Db      1935 CCATCTTTAACCTTAGGGAAATTGACAATTATGATGAGTATGAAAGATCAATGTTGATCT 1994
```

```
Qy      1900 CCCAGTTGAGCTTTGAGAAAACTTTTGGCTTATCTTCTGTCTTCATTGAGTCTACACTAA 1959
Db      1995 CTCAAACAAGCTTTGAGAAAACTTTTGGCTTATCTTCAGTCTTCATTGAATCTACACTAA 2054

Qy      1960 TGGAGAATGGAGGAGTACCCGAGTCTGCCAACTCACCACCATTCATCAAGGAAGCGATTC 2019
Db      2055 TGGAGAATGGAGGAGTGGCTGAATCTGCCAACCCTTCCACACTAATCAAGGAAGCAATTC 2114

Qy      2020 AAGTCATCGGCTGTGGCTATGAAGAGAAGACTGAATGGGGAAAACAGATTGGTTGGATAT 2079
Db      2115 ATGTCATCAGCTGTGGCTATGAAGAGAAGACTGCATGGGGGAAAGAGATTGGATGGATAT 2174

Qy      2080 ATGGGTCAGTCACTGAGGATATCTTAAGTGGCTTCAAGATGCACTGCCGAGGATGGAGAT 2139
Db      2175 ATGGTTCAGTCACTGAGGATATCTTAACCGGCTTCAAAATGCACTGCCGAGGATGGAGAT 2234

Qy      2140 CAATTTACTGCATGCCCGTAAGGCCTGCATTCAAAGGATCTGCACCCATCAACCTGTCTG 2199
Db      2235 CGATTTACTGCATGCCCTTAAGGCCAGCATTCAAAGGATCTGCACCCATCAATCTGTCTG 2294

Qy      2200 ATAGATTGCACCAGGTCCTCCGATGGGCTCTTGGTTCTGTGGAAATTTTCTTTAGCAGAC 2259
Db      2295 ATCGGTTGCACCAGGTTCTTCGATGGGCTCTTGGATCTGTTGAAATTTTCCTAAGCAGGC 2354

Qy      2260 ACTGTCCCCTCTGGTACGGGTTTGGAGGAGGCCGTCTTAAATGGCTCCAAAGGCTTGCGT 2319
Db      2355 ATTGCCCTCTATGGTATGGCTTTGGAGGTGGTCGTCTTAAATGGCTTCAAAGACTAGCAT 2414

Qy      2320 ATATAAACACCATTGTGTACCCATTTACATCCCTCCCTCTCATTGCCTATTGCACAATTC 2379
Db      2415 ATATAAACACCATTGTCTATCCTTTCACATCCCTTCCACTCATTGCCTATTGTTCACTAC 2474

Qy      2380 CTGCAGTTTGTCTGCTCACCGGAAAATTCATCATACCAACGCTCTCAAACCTGGCAAGCA 2439
Db      2475 CAGCAATCTGTCTTCTCACAGGAAAATTTATCATACCAACGCTCTCAAACCTGGCAAGTG 2534

Qy      2440 TGCTGTTTCTTGGCCTCTTTATCTCCATCATTGTAACTGCGGTGCTTGAGCTAAGATGGA 2499
Db      2535 TTCTCTTTCTTGGCCTTTTCCTTTCCATTATCGTGACTGCTGTTCTCGAGCTCCGATGGA 2594

Qy      2500 GCGGTGTCAGCATTGAAGATTTATGGCGTAATGAACAATTCTGGGTGATCGGAGGTGTTT 2559
Db      2595 GTGGTGTCAGCATTGAGGACTTATGGCGTAACGAGCAGTTTTGGGTCATCGGTGGCGTTT 2654

Qy      2560 CAGCCCATCTCTTTGCGGTCTTCCAGGGATTCTTAAAAATGTTGGCTGGCATCGATACGA 2619
Db      2655 CAGCCCATCTCTTTGCCGTCTTCCAAGGTTTCCTTAACATGTTGCCGGCATTGACACCA 2714

Qy      2620 ACTTCACTGTCACAGCAAAAGCAGCTGATGATCAGATTTCGGAGCTCTACATTGTGA 2679
Db      2715 ACTTTACTGTCACTGCCAAAGCAACTGATGATCAGATTTCGGAGCTCTACATTGTGA 2774

Qy      2680 AGTGGACAACACTTTTGATTCCTCCAACCACACTTCTCATTATCAATATGTCGGGT---T 2736
Db      2775 AATGGACTACACTTCTAATCCCTCCAACAACACTCCTCATCGTCAACATGGTTGGTGTCG 2834

Qy      2737 GTGCTGGATTCTCTGATGCACTCAACAAAGGATATGAAGCATGGGGGCCTCTCTTTGGCA 2796
Db      2835 TTGCCGGATTCTCCGATGCCCTCAACAAAGGGTATGAAGCTTGGGGACCACTCTTTGGCA 2894

Qy      2797 AGGTGTTCTTTGCTTTCTGGGTGATTCTTCATCTCTATCCATTCCTTAAAGGTCTAATGG 2856
Db      2895 AAGTGTTCTTTTCCTTCTGGGTCATCCTCCATCTTTATCCATTCCTCAAAGGTCTTATGG 2954

Qy      2857 GTCGCCAAAACCTAACACCAACCATTGTTGTTCTCTGGTCAGTGCTGTTGGCCTCTGTCT 2916
Db      2955 GACGCCAAAACAGGACACCAACCATTGTTGTCCTTTGGTCAGTGTTGTTGGCTTCTGTCT 3014

Qy      2917 TCTCTCTCGTTTGGGTCAAGATCAATCCATTCGTTAACAAAGTTGATAACACCTTGGTTG 2976
Db      3015 TCTCTCTTGTTTGGGTTCGGATCAACCCGTTTGTCAGCACCGCCGATAGCACCACCGTGT 3074

Qy      2977 CGGAGACCTGCATTTCCATTGATTGCTGA 3006
```

```
Db      3075 CACAGAGCTGCATTTCCATTGATTGTTGA 3103
```

RESULT 3
AAV08372
ID AAV08372 standard; cDNA to mRNA; 3207 BP.
XX
AC AAV08372;
XX
DT 08-FEB-1999 (first entry)
XX
DE Cellulose synthase PcsA1 coding sequence.
XX
KW Cellulose synthase; cotton; ds.
XX
OS Gossypium hirsutum.
XX
FH Key   Location/Qualifiers
FT CDS   77..3001
FT   /*tag= a
XX
PN EP875575-A2.
XX
PD 04-NOV-1998.
XX
PF 31-MAR-1998; 98EP-0302489.
XX
PR 01-APR-1997; 97JP-0083133.
XX
PA (NISN) NISSHINBO IND INC.
XX
PI Aotsuka S, Hasegawa O, Hayashi T, Ihara Y;
XX
DR WPI; 1998-559440/48.
DR P-PSDB; AAW73308.
XX
PT New DNA encoding cotton cellulose synthase proteins - useful for
PT controlling cellulose synthesis in prokaryotic or eukaryotic cells
XX
PS Example 5; Page 14-18; 51pp; English.
XX
CC This sequence encodes a cellulose synthase of the invention. The DNA can
CC be used for controlling cellulose synthesis in a cell by introducing the
CC DNA into the cell, expressing RNA with a nucleotide sequence homologous
CC to the DNA or a nucleotide sequence complementary to the DNA. The cells
CC are prokaryotic cells i.e. acetobacterium and/or eukaryotic cells such as
CC yeasts i.e. Saccharomyces spp., plant cells (cotton plants) and cultured
CC cells of mammals etc.
XX
SQ Sequence 3207 BP; 883 A; 662 C; 723 G; 939 T; 0 other;

Query Match 61.5%; Score 1988.2; DB 19; Length 3207;
Best Local Similarity 80.8%; Pred. No. 0;
Matches 2398; Conservative 0; Mismatches 553; Indels 18; Gaps 6;

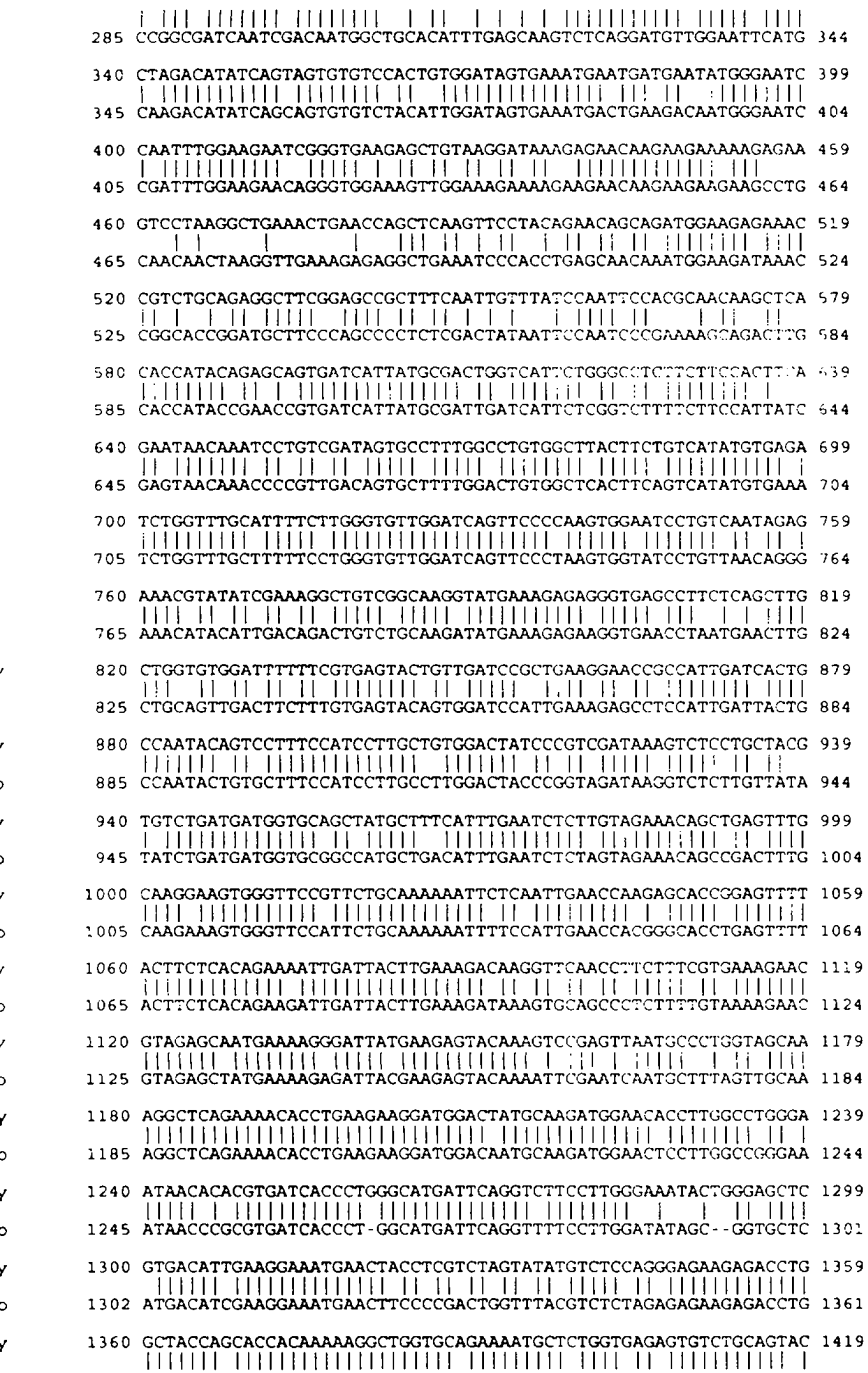
```
Qy        40 AAAGAGCTGGTAAAGTGGTAATAAGCAAGATGATGGAATCTGGGGCTCCTATATGCCATA 99
Db        48 AAGGAGAAAGATAAGTAATGTTTTTGAGAATGATGGAATCTGGGGTTCCTGTTTGCCACA 107

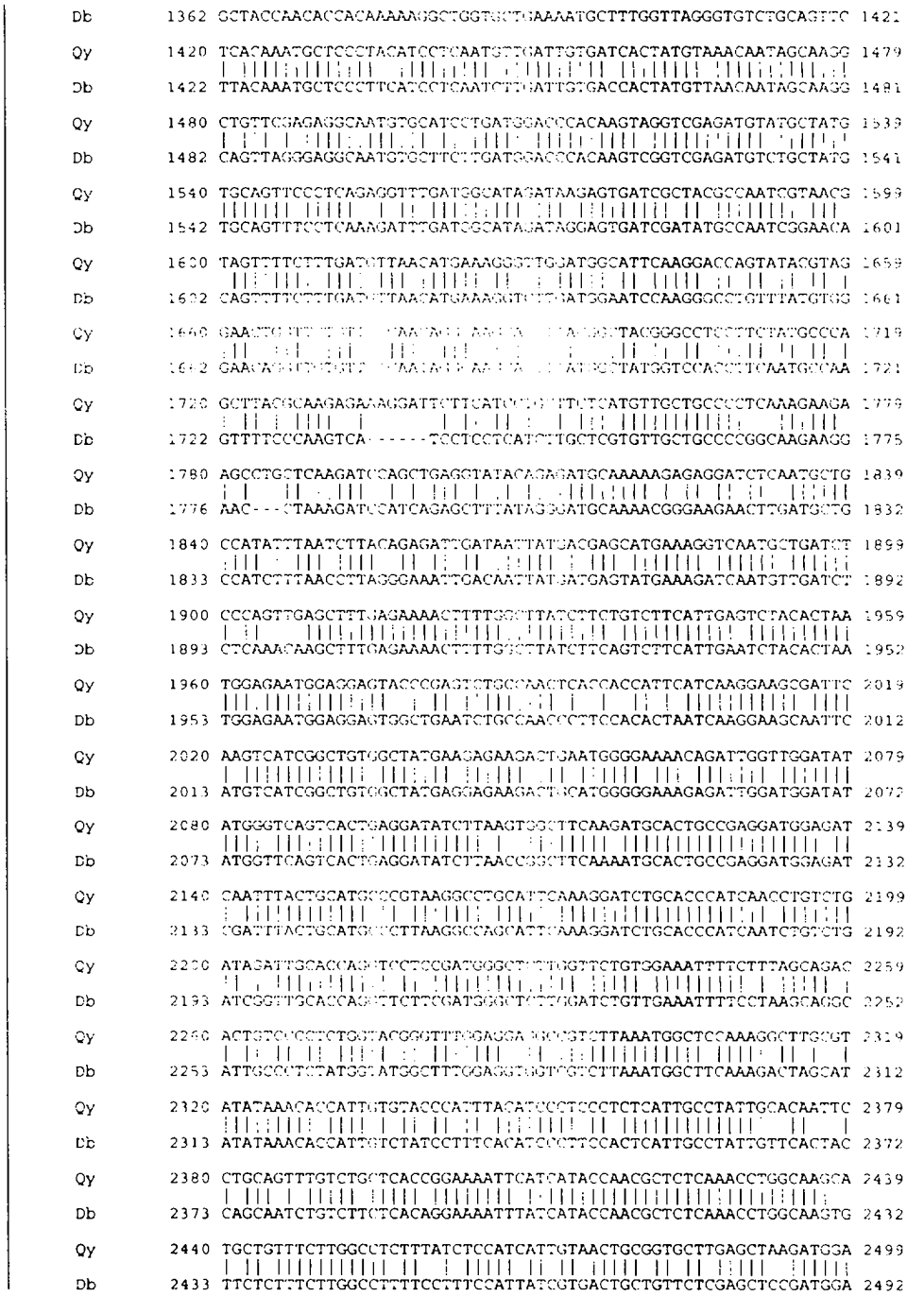
Qy       100 CCTGTGGTGAACAGGTGGGGCATGATGCAAATGGGGAGCTATTTGTGGCTTGCCATGAGT 159
Db       108 CTTGTGGTGAACATGTTGGGTTGAATGTTAATGGTGAACCCTTTGTGGCTTGCCATGAAT 167

Qy       160 GTAGCTATCCCATGTGCAAGTCTTGTTTCGAGTTTGAAATCAATGAGGGCCGGAAAGTTT 219
Db       168 GTAATTTCCCTATTTGTAAGAGTTGTTTTGAGTATGATCTTAAGGAAGGACAAAAAGCTT 227

Qy       220 GCTTGCGGTGTGGCTCGCCATATGATGAGAACTTGCTGGATGATGTAGAAAAGAAGGGGT 279
Db       228 GCTTGCGTTGTGGTATTCCGTATGATGAAAACCTGTTGGACGATGTCG---AGAAGGCCA 284

Qy       280 CTGGCAATCAATCCACAATGGCATCTCACCTCAACGATTCTCAGGATGTCGGAATCCATG 339
```


285 CCGGCGATCAATCGACAATGGCTGCACATTTGAGCAAGTCTCAGGATGTTGGAATTCATG 344
340 CTAGACATATCAGTAGTGTGTCCACTGTGGATAGTGAAATGAATGATGAATATGGGAATC 399
345 CAAGACATATCAGCAGTGTGTCTACATTGGATAGTGAAATGACTGAAGACAATGGGAATC 404
400 CAATTTGGAAGAATCGGGTGAAGAGCTGTAAGGATAAAGAGAACAAGAAGAAAAAGAGAA 459
405 CGATTTGGAAGAACAGGGTGGAAAGTTGGAAAGAAAAGAAGAACAAGAAGAAGAAGCCTG 464
460 GTCCTAAGGCTGAAACTGAACCAGCTCAAGTTCCTACAGAACAGCAGATGGAAGAGAAAC 519
465 CAACAACTAAGGTTGAAAGAGAGGCTGAAATCCCACCTGAGCAACAAATGGAAGATAAAC 524
520 CGTCTGCAGAGGCTTCGGAGCCGCTTTCAATTGTTTATCCAATTCCACGCAACAAGCTCA 579
525 CGGCACCGGATGCTTCCCAGCCCCTCTCGACTATAATTCCAATCCCGAAAAGCAGACTTG 584
580 CACCATACAGAGCAGTGATCATTATGCGACTGGTCATTCTGGGCCTCTTCTTCCACTTCA 639
585 CACCATACCGAACCGTGATCATTATGCGATTGATCATTCTCGGTCTTTTCTTCCATTATC 644
640 GAATAACAAATCCTGTCGATAGTGCCTTTGGCCTGTGGCTTACTTCTGTCATATGTGAGA 699
645 GAGTAACAAACCCCGTTGACAGTGCTTTTGGACTGTGGCTCACTTCAGTCATATGTGAAA 704
700 TCTGGTTTGCATTTTCTTGGGTGTTGGATCAGTTCCCCAAGTGGAATCCTGTCAATAGAG 759
705 TCTGGTTTGCTTTTTCCTGGGTGTTGGATCAGTTCCCTAAGTGGTATCCTGTTAACAGGG 764
760 AAACGTATATCGAAAGGCTGTCGGCAAGGTATGAAAGAGAGGGTGAGCCTTCTCAGCTTG 819
765 AAACATACATTGACAGACTGTCTGCAAGATATGAAAGAGAAGGTGAACCTAATGAACTTG 824
820 CTGGTGTGGATTTTTTCGTGAGTACTGTTGATCCGCTGAAGGAACCGCCATTGATCACTG 879
825 CTGCAGTTGACTTCTTTGTGAGTACAGTGGATCCATTGAAAGAGCCTCCATTGATTACTG 884
880 CCAATACAGTCCTTTCCATCCTTGCTGTGGACTATCCCGTCGATAAAGTCTCCTGCTACG 939
885 CCAATACTGTGCTTTCCATCCTTGCCTTGGACTACCCGGTAGATAAGGTCTCTTGTTATA 944
940 TGTCTGATGATGGTGCAGCTATGCTTTCATTTGAATCTCTTGTAGAAACAGCTGAGTTTG 999
945 TATCTGATGATGGTGCGGCCATGCTGACATTTGAATCTCTAGTAGAAACAGCCGACTTTG 1004
1000 CAAGGAAGTGGGTTCCGTTCTGCAAAAAATTCTCAATTGAACCAAGAGCACCGGAGTTTT 1059
1005 CAAGAAAGTGGGTTCCATTCTGCAAAAAATTTTCCATTGAACCACGGGCACCTGAGTTTT 1064
1060 ACTTCTCACAGAAAATTGATTACTTGAAAGACAAGGTTCAACCTTCTTTCGTGAAAGAAC 1119
1065 ACTTCTCACAGAAGATTGATTACTTGAAAGATAAAGTGCAGCCCTCTTTTGTAAAAGAAC 1124
1120 GTAGAGCAATGAAAAGGGATTATGAAGAGTACAAAGTCCGAGTTAATGCCCTGGTAGCAA 1179
1125 GTAGAGCTATGAAAAGAGATTACGAAGAGTACAAAATTCGAATCAATGCTTTAGTTGCAA 1184
1180 AGGCTCAGAAAACACCTGAAGAAGGATGGACTATGCAAGATGGAACACCTTGGCCTGGGA 1239
1185 AGGCTCAGAAAACACCTGAAGAAGGATGGACAATGCAAGATGGAACTCCTTGGCCGGGAA 1244
1240 ATAACACACGTGATCACCCTGGGCATGATTCAGGTCTTCCTTGGGAAATACTGGGAGCTC 1299
1245 ATAACCCGCGTGATCACCCT-GGCATGATTCAGGTTTTCCTTGGATATAGC--GGTGCTC 1301
1300 GTGACATTGAAGGAAATGAACTACCTCGTCTAGTATATGTCTCCAGGGAGAAGAGACCTG 1359
1302 ATGACATCGAAGGAAATGAACTTCCCCGACTGGTTTACGTCTCTAGAGAGAAGAGACCTG 1361
1360 GCTACCAGCACCACAAAAAGGCTGGTGCAGAAAATGCTCTGGTGAGAGTGTCTGCAGTAC 1419


Db 1362 GCTACCAACACCACAAAAAGGCTGGTGCTGAAAATGCTTTGGTTAGGGTGTCTGCAGTTC 1421
Qy 1420 TCACAAATGCTCCCTACATCCTCAATGTTGATTGTGATCACTATGTAAACAATAGCAAGG 1479
Db 1422 TTACAAATGCTCCCTTCATCCTCAATCTTGATTGTGACCACTATGTTAACAATAGCAAGG 1481
Qy 1480 CTGTTCGAGAGGCAATGTGCATCCTGATGGACCCACAAGTAGGTCGAGATGTATGCTATG 1539
Db 1482 CAGTTAGGGAGGCAATGTGCTTCCTGATGGACCCACAAGTCGGTCGAGATGTCTGCTATG 1541
Qy 1540 TGCAGTTCCCTCAGAGGTTTGATGGCATAGATAAGAGTGATCGCTACGCCAATCGTAACG 1599
Db 1542 TGCAGTTTCCTCAAAGATTTGATCGCATAGATAGGAGTGATCGATATGCCAATCGGAACA 1601
Qy 1600 TAGTTTTCTTTGATGTTAACATGAAAGGGTTGGATGGCATTCAAGGACCAGTATACGTAG 1659
Db 1602 CAGTTTTCTTTGATGTTAACATGAAAGGTTTTGATGGAATCCAAGGGCCTGTTTATGTGG 1661
Qy 1660 GAACTGTTTTTTCTTCAAAGAGAAAGTACTACGGCCTGCTTCTATGCCCA 1719
Db 1662 GAACAGTTTTTGTTCAAAGATAAATTACTATGGTCCACCTTCAATGCCAA 1721
Qy 1720 GCTTACGCAAGAGAAAGGATTCTTGATCCGTTTCTCATGTTGCTGCCCCTCAAAGAAGA 1779
Db 1722 GTTTTCCCAAGTCA-----TCCTCCTCATCTTGCTCGTGTTGCTGCCCCGGCAAGAAGG 1775
Qy 1780 AGCCTGCTCAAGATCGAGCTGAGGTATACAGAGATGCAAAAAGAGAGGATCTCAATGCTG 1839
Db 1776 AAC---CTAAAGATCCATCAGAGCTTTATAGGGATGCAAAACGGGAAGAACTTGATGCTG 1832
Qy 1840 CCATATTTAATCTTACAGAGATTGATAATTATGACGAGCATGAAAGGTCAATGCTGATCT 1899
Db 1833 CCATCTTTAACCTTAGGGAAATTGACAATTATGATGAGTATGAAAGATCAATGTTGATCT 1892
Qy 1900 CCCAGTTGAGCTTTGAGAAAACTTTTGGCTTATCTTCTGTCTTCATTGAGTCTACACTAA 1959
Db 1893 CTCAAACAAGCTTTGAGAAAACTTTTGGCTTATCTTCAGTCTTCATTGAATCTACACTAA 1952
Qy 1960 TGGAGAATGGAGGAGTACCCGAGTCTGCCAACTCACCACCATTCATCAAGGAAGCGATTC 2019
Db 1953 TGGAGAATGGAGGAGTGGCTGAATCTGCCAACCCTTCCACACTAATCAAGGAAGCAATTC 2012
Qy 2020 AAGTCATCGGCTGTGGCTATGAAGAGAAGACTGAATGGGGAAAACAGATTGGTTGGATAT 2079
Db 2013 ATGTCATCGGCTGTGGCTATGAGGAGAAGACTGCATGGGGGAAAGAGATTGGATGGATAT 2072
Qy 2080 ATGGGTCAGTCACTGAGGATATCTTAAGTGGCTTCAAGATGCACTGCCGAGGATGGAGAT 2139
Db 2073 ATGGTTCAGTCACTGAGGATATCTTAACCGGCTTCAAAATGCACTGCCGAGGATGGAGAT 2132
Qy 2140 CAATTTACTGCATGCCCGTAAGGCCTGCATTCAAAGGATCTGCACCCATCAACCTGTCTG 2199
Db 2133 CGATTTACTGCATGCCCTTAAGGCCAGCATTCAAAGGATCTGCACCCATCAATCTGTCTG 2192
Qy 2200 ATAGATTGCACCAGATCCTCCGATGGGCTCTTGGTTCTGTGGAAATTTTCTTTAGCAGAC 2259
Db 2193 ATCGGTTGCACCAGATCTTCTCGATGGGCTCTTGGATCTGTTGAAATTTTCCTAAGCAGGC 2252
Qy 2260 ACTGTCCCGTCTGGTACGGGTTTGGAGGATGCCGTCTTAAATGGCTCCAAAGGCTTGCGT 2319
Db 2253 ATTGCCCTCTATGGCATGGCTTTGGAGGTGGTCGTCTTAAATGGCTTCAAAGACTAGCAT 2312
Qy 2320 ATATAAACACCATTGTGTACCCATTTACATCCCTCCCTCTCATTGCCTATTGCACAATTC 2379
Db 2313 ATATAAACACCATTGTCTATCCTTTCACATCCCTTCCACTCATTGCCTATTGTTCACTAC 2372
Qy 2380 CTGCAGTTTGTCTGCTCACCGGAAAATTCATCATACCAACGCTCTCAAACCTGGCAAGCA 2439
Db 2373 CAGCAATCTGTCTTCTCACAGGAAAATTTATCATACCAACGCTCTCAAACCTGGCAAGTG 2432
Qy 2440 TGCTGTTTCTTGGCCTCTTTATCTCCATCATTGTAACTGCGGTGCTTGAGCTAAGATGGA 2499
Db 2433 TTCTCTTTCTTGGCCTTTTCCTTTCCATTATCGTGACTGCTGTTCTCGAGCTCCGATGGA 2492

```
Qy       2500 GCGGTGTCAGCATTGAAGATTTATGGCGTAATGAACAATTCTGGGTGATCGGAGGTGTTT 2559
Db       2493 GTGGTGTCAGCATTGAGGACTTATGGCGTAACGAGCAGTTTTGGGTCATCGGTGGCGTTT 2552

Qy       2560 CAGCCCATCTCTTTGCGGTCTTCCAGGGATTCTTAAAAATGTTGGCTGGCATCGATACGA 2619
Db       2553 CAGCCCATCTCTTTGCCGTCTTCCAAGGTTTCCTTAAGATGCTTGCGGGCATTGACACCA 2612

Qy       2620 ACTTCACTGTCACAGCAAAAGCAGCCGAAGATGCAGAATTTGGGGAGCTATATATGGTCA 2679
Db       2613 ACTTTACTGTCACTGCCAAAGCAGCTGATGATGCAGATTTTGGTGAGCTCTACATTGTGA 2672

Qy       2680 AGTGGACAACACTTTTGATTCCTCCAACCACACTTCTCATTATCAATATGTCGGGT---T 2736
Db       2673 AATGGACTACACTTCTAATCCCTCCAACAACACTCCTCATCGTCAACATGGTTGGTGTCG 2732

Qy       2737 GTGCTGGATTCTCTGATGCACTCAACAAAGGATATGAAGCATGGGGGCCTCTCTTTGGCA 2796
Db       2733 TTGCCGGATTCTCCGATGCCCTCAACAAAGGGTACGAAGCTTGGGGACCACTCTTTGGCA 2792

Qy       2797 AGGTGTTCTTTGCTTTCTGGGTGATTCTTCATCTCTATCCATTCCTTAAAGGTCTAATGG 2856
Db       2793 AAGTGTTCTTTTCCTTCTGGGTCATCCTCCATCTTTATCCATTCCTCAAAGGTCTTATGG 2852

Qy       2857 GTCGCCAAAACCTAACACCAACCATTGTTGTTCTCTGGTCAGTGCTGTTGGCCTCTGTCT 2916
Db       2853 GACGCCAAAACAGGACACCAACCATTGTTGTCCTTGGTCAGTGTTGTTGGCTTCTGTCT 2912

Qy       2917 TCTCTCTCGTTTGGGTCAAGATCAATCCATTCGTTAACAAAGTTGATAACACCTTGGTTG 2976
Db       2913 TCTCTCTTGTTTGGGTTCGGATCAACCCGTTTGTCAGCACCGCCGATAGCACCACCGTGT 2972

Qy       2977 CGGAGACCTGCATTTCCATTGATTGCTGA 3005
Db       2973 CACAGAGCTGCATTTCCATTGATTGTTGA 3001
```

RESULT 4
AAA67145
ID AAA67145 standard; DNA; 3747 BP.
XX
AC AAA67145;
XX
DT 31-OCT-2000 (first entry)
XX
DE Eucalyptus grandis cellulose synthase nucleotide sequence SEQ ID NO:141.
XX
KW Eucalyptus grandis; pinus radiata; Monterey pine; modification;
KW plant cell wall; polysaccharide; polysaccharide biosynthetic pathway;
KW transgenic plant; ds.
XX
OS Eucalyptus grandis.
XX
PN WO200022092-A2.
XX
PD 20-APR-2000.
XX
PF 08-OCT-1999; 99WO-NZ00169.
XX
PR 13-OCT-1998; 98US-0170862.
PR 11-AUG-1999; 99US-0148426.
XX
PA (GENE-) GENESIS RES & DEV CORP LTD.
PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI Bicksberg LN;
XX
DR WPI; 2000-339328/29.
DR P-PSDB; AAB16338.
XX
PT New genes encoding proteins involved in a plant polysaccharide
PT biosynthetic pathway, useful for modulating or altering the
PT polysaccharide content, composition or structure of the plant -
XX
PS Claim 1; Page 106-107; 301pp; English.
XX
CC The present invention describes isolated polynucleotides (PN) comprising
CC a sequence selected from one of 835 nucleotide sequences given in
CC AAA67073 to AAA67907, their (reverse) complements, sequences producing
CC an Expectation (E) value of 0.01 or less compared to the 835 sequences,
CC sequences at least 50% identical to them, 200, 100, 40 or 20-mers of the
CC 835 sequences or sequences that are degenerately equivalent or allelic
CC to the 835 sequences. The polynucleotides are used to modify the
CC activity of a polypeptide involved in a polysaccharide biosynthetic
CC pathway in the plant. They are especially used to modulate or alter the
CC polysaccharide content, composition or structure of the plant. AAB16268
CC to AAB16340 are proteins encoded by some of the polynucleotide sequence
CC given in the present invention.
XX
SQ Sequence 3747 BP; 976 A; 822 C; 935 G; 1014 T; 0 other;

  Query Match   57.2%; Score 1848.6; DB 21; Length 3747;
  Best Local Similarity   78.2%; Pred. No. 0;
  Matches 2312; Conservative   0; Mismatches 624; Indels   19; Gaps   7;

```
Qy         65 CAAGATGATGGAATCTGGGGCTCCTATATGCCATACCTGTGGTGAACAGGTGGGGCATGA 124
Db        507 CAACATGATGGAATCCGGGGTTCCCCTGTGCAACACTTGCGGAGAGGCTGTTGGGGTTGA 566

Qy        125 TGCAAATGGGGAGCTATTTGTGGCTTGCCATGAGTGTAGCTATCCCATGTGCAAGTCTTG 184
Db        567 TGAGAAAGGCGAGGTCTTCGTGGCTTGTCAAGAGTGCAACTTCGCCATTTGCAAGGCTTG 626

Qy        185 TTTCGAGTTTGAAATCAATGAGGGCCGGAAAGTTTGCTTGCGGTGTGGCTCGCCATATGA 244
Db        627 TGTCGAATATGAGATTAAGGAAGGAAGAAAAGCGTGCTTGCGCTGTGGCACTCCATTTGA 686

Qy        245 TGAGAACTTGCTGGATGATGTAGAAAAGAAGGGGTCTGGCAATCAATCCACAATGGCATC 304
Db        687 AGCGAACTCGATGGCTGATGCTGAGAGAAATGAATTGGGAAGTCGATCGACAATGGCAGC 746

Qy        305 TCACCTCAACGATTCTCAGGATGTCGGAATCCATGCTAGACATATCAGTAGTGTGTCCAC 364
Db        747 TCAACTCAATGATCCTCAGGACACAGGGATTCATGCTAGACACATCAGCAGTGTTTCTAC 806

Qy        365 TGTGGATAGTGAAATGAATGATGAATATGGGAATCCAATTTGGAAGAATCGGGTGAAGAG 424
Db        807 GTTGGATAGTGAATACAATGATGAGACTGGGAACCCTATCTGGAAGAATAGAGTGGAGAG 866

Qy        425 CTGTAAGGATAAAGAGAACAAGAAGAAAAAGAGAAGTCCTAAGGCTGAAACTGAACCAGC 484
Db        867 CTGGAAGGACAAAAAGAATAAGAAGAAGAAGGCCCCGACGAAGGCTGAGA---AAGAGGC 923

Qy        485 TCAAGTTCCTACAGAACAGCAGATGGAAGAGAAACCGTCTGCAGAGGCTTCGGAGCCGCT 544
Db        924 TCAAGTTCCACCAGAGCAGCAGATGGAAGAGAAGCAAATTGCTGATGCTTCAGAGCCACT 983

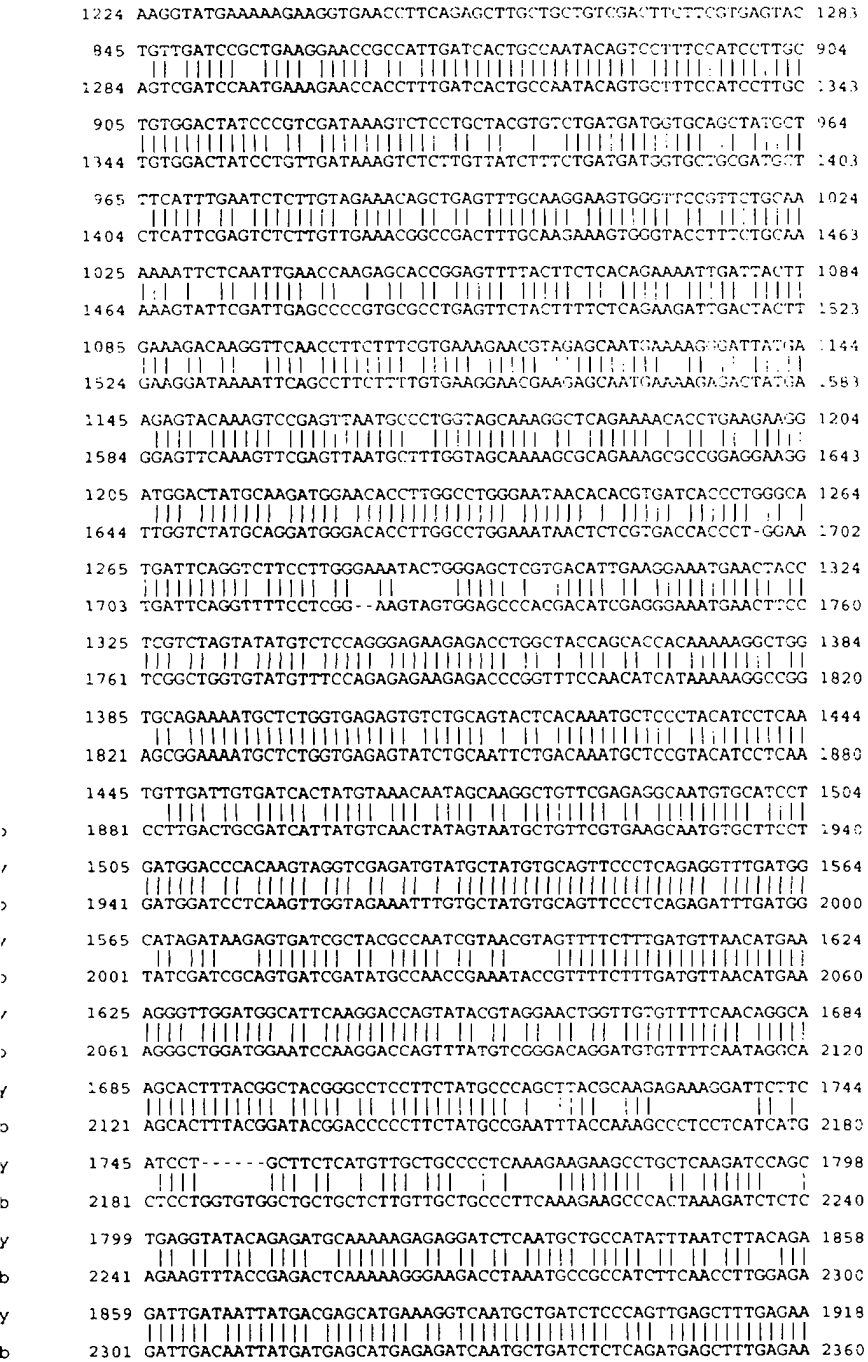
Qy        545 TTCAATTGTTTATCCAATTCCACGCAACAAGCTCACACCATACAGAGCAGTGATCATTAT 604
Db        984 CTCGACCGTTATTCCCATTGCCAAAAGCAAACTCGCACCATACCGAACCGTAATAATAAT 1043

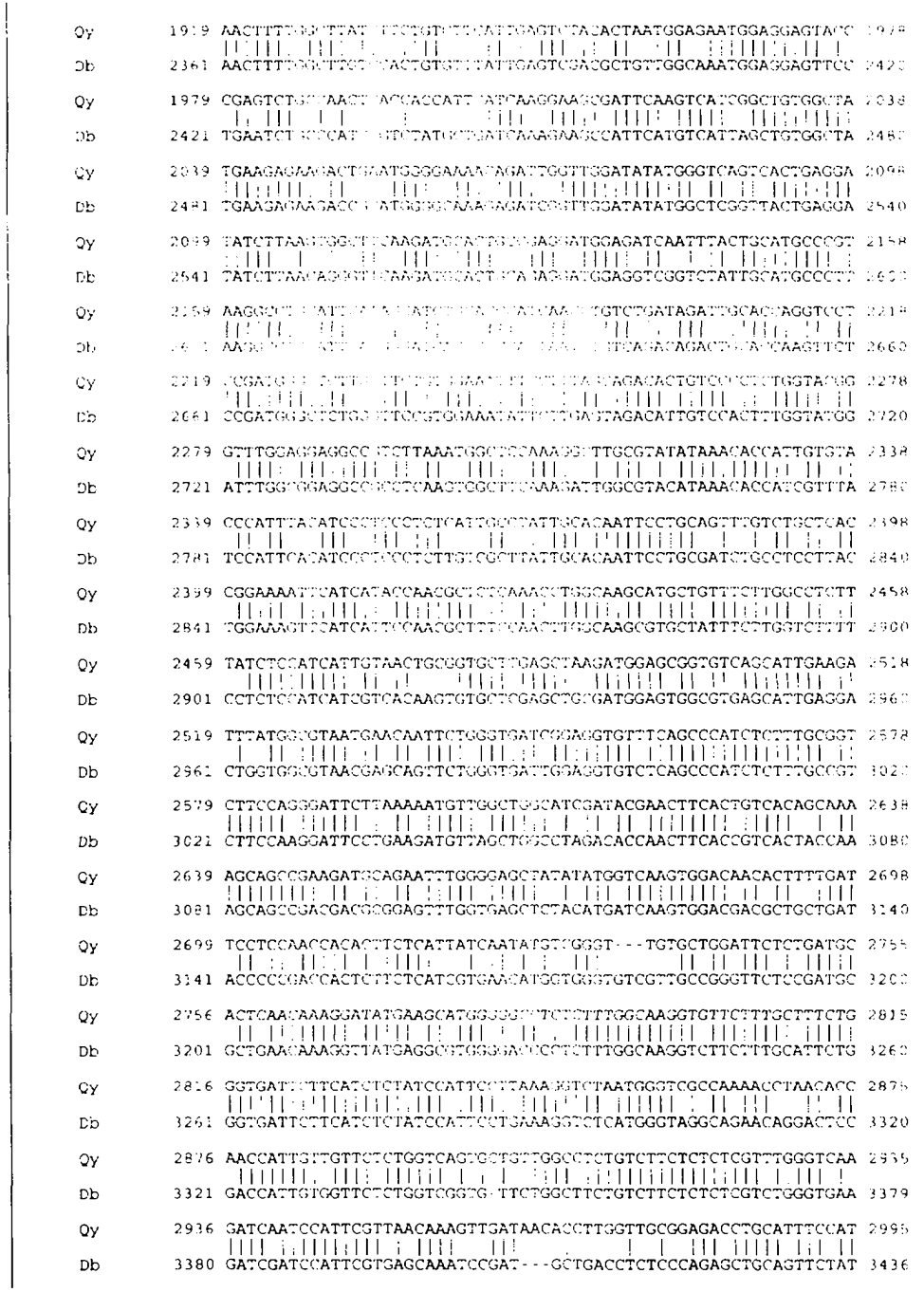
Qy        605 GCGACTGGTCATTCTGGGCCTCTTCTTCCACTTCAGAATAACAAATCCTGTCGATAGTGC 664
Db       1044 GCGATTGATCATTTTGGCACTTTTCTTCCATTACCGAGTGACACATCCAGTCGACAGCGC 1103

Qy        665 CTTTGGCCTGTGGCTTACTTCTGTCATATGTGAGATCTGGTTTGCATTTTCTTGGGTGTT 724
Db       1104 CTATCCATTGTGGCTTACATCCATCATATGTGAGATTTGGTTCGCATACTCATGGGTGTT 1163

Qy        725 GGATCAGTTCCCCAAGTGGAATCCTGTCAATAGAGAAACGTATATCGAAAGGCTGTCGGC 784
Db       1164 GGATCAGTTCCCGAAGTGGTCTCCCGTGAATAGGATTACTCATGTTGACAGACTATCTGC 1223

Qy        785 AAGGTATGAAAGAGAGGGTGAGCCTTCTCAGCTTGCTGGTGTGGATTTTTTCGTGAGTAC 844
```


1224 AAGGTATGAAAAAGAAGGTGAACCTTCAGAGCTTGCTGCTGTCGACTTCTTCGTGAGTAC 1283
845 TGTTGATCCGCTGAAGGAACCGCCATTGATCACTGCCAATACAGTCCTTTCCATCCTTGC 904
1284 AGTCGATCCAATGAAAGAACCACCTTTGATCACTGCCAATACAGTGCTTTCCATCCTTGC 1343
905 TGTGGACTATCCCGTCGATAAAGTCTCCTGCTACGTGTCTGATGATGGTGCAGCTATGCT 964
1344 TGTGGACTATCCTGTTGATAAAGTCTCTTGTTATCTTTCTGATGATGGTGCTGCGATGCT 1403
965 TTCATTTGAATCTCTTGTAGAAACAGCTGAGTTTGCAAGGAAGTGGGTTCCGTTCTGCAA 1024
1404 CTCATTCGAGTCTCTTGTTGAAACGGCCGACTTTGCAAGAAAGTGGGTACCTTTCTGCAA 1463
1025 AAAATTCTCAATTGAACCAAGAGCACCGGAGTTTTACTTCTCACAGAAAATTGATTACTT 1084
1464 AAAGTATTCGATTGAGCCCCGTGCGCCTGAGTTCTACTTTTCTCAGAAGATTGACTACTT 1523
1085 GAAAGACAAGGTTCAACCTTCTTTCGTGAAAGAACGTAGAGCAATGAAAAGTGATTATGA 1144
1524 GAAGGATAAAATTCAGCCTTCTTTTGTGAAGGAACGAAGAGCAATGAAAAGAGACTATGA 1583
1145 AGAGTACAAAGTCCGAGTTAATGCCCTGGTAGCAAAGGCTCAGAAAACACCTGAAGAAGG 1204
1584 GGAGTTCAAAGTTCGAGTTAATGCTTTGGTAGCAAAAGCGCAGAAAGCGCCGGAGGAAGG 1643
1205 ATGGACTATGCAAGATGGAACACCTTGGCCTGGGAATAACACACGTGATCACCCTGGGCA 1264
1644 TTGGTCTATGCAGGATGGGACACCTTGGCCTGGAAATAACTCTCGTGACCACCCT-GGAA 1702
1265 TGATTCAGGTCTTCCTTGGGAAATACTGGGAGCTCGTGACATTGAAGGAAATGAACTACC 1324
1703 TGATTCAGGTTTTCCTCGG--AAGTAGTGGAGCCCACGACATCGAGGGAAATGAACTTCC 1760
1325 TCGTCTAGTATATGTCTCCAGGGAGAAGAGACCTGGCTACCAGCACCACAAAAAGGCTGG 1384
1761 TCGGCTGGTGTATGTTTCCAGAGAGAAGAGACCCGGTTTCCAACATCATAAAAAGGCCGG 1820
1385 TGCAGAAAATGCTCTGGTGAGAGTGTCTGCAGTACTCACAAATGCTCCCTACATCCTCAA 1444
1821 AGCGGAAAATGCTCTGGTGAGAGTATCTGCAATTCTGACAAATGCTCCGTACATCCTCAA 1880
1445 TGTTGATTGTGATCACTATGTAAACAATAGCAAGGCTGTTCGAGAGGCAATGTGCATCCT 1504
1881 CCTTGACTGCGATCATTATGTCAACTATAGTAATGCTGTTCGTGAAGCAATGTGCTTCCT 1940
1505 GATGGACCCACAAGTAGGTCGAGATGTATGCTATGTGCAGTTCCCTCAGAGGTTTGATGG 1564
1941 GATGGATCCTCAAGTTGGTAGAAATTTGTGCTATGTGCAGTTCCCTCAGAGATTTGATGG 2000
1565 CATAGATAAGAGTGATCGCTACGCCAATCGTAACGTAGTTTTCTTTGATGTTAACATGAA 1624
2001 TATCGATCGCAGTGATCGATATGCCAACCGAAATACCGTTTTCTTTGATGTTAACATGAA 2060
1625 AGGGTTGGATGGCATTCAAGGACCAGTATACGTAGGAACTGGTTGTGTTTTCAACAGGCA 1684
2061 AGGGCTGGATGGAATCCAAGGACCAGTTTATGTCGGGACAGGATGTGTTTTCAATAGGCA 2120
1685 AGCACTTTACGGCTACGGGCCTCCTTCTATGCCCAGCTTACGCAAGAGAAAGGATTCTTC 1744
2121 AGCACTTTACGGATACGGACCCCCTTCTATGCCGAATTTACCAAAGCCCTCCTCATCATG 2180
1745 ATCCT------GCTTCTCATGTTGCTGCCCCCTCAAAGAAGAAGCCTGCTCAAGATCCAGC 1798
2181 CTCCTGGTGTGGCTGCTGCTCTTGTTGCTGCCCTTCAAAGAAGCCCACTAAAGATCTCTC 2240
1799 TGAGGTATACAGAGATGCAAAAAGAGAGGATCTCAATGCTGCCATATTTAATCTTACAGA 1858
2241 AGAAGTTTACCGAGACTCAAAAAGGGAAGACCTAAATGCCGCCATCTTCAACCTTGGAGA 2300
1859 GATTGATAATTATGACGAGCATGAAAGGTCAATGCTGATCTCCCAGTTGAGCTTTGAGAA 1918
2301 GATTGACAATTATGATGAGCATGAGAGATCAATGCTGATCTCTCAGATGAGCTTTGAGAA 2360


Qy 1919 AACTTTTGGCTTAT TCCTGTTCAATTGAGTCTACACTAATGGAGAATGGAGGAGTACC 1978
Db 2361 AACTTTTGGCTTCT TACTGTGTTTATTGAGTCGACGCTGTTGGCAAATGGAGGAGTTCC 2420
Qy 1979 CGAGTCTGCTAACT ACCACCATT ATCAAGGAAGCGATTCAAGTCA CGGCTGTGGCTA 2038
Db 2421 TGAATCT CCCAT TGCTATGCTGATCAAAGAAGCCATTCATGTCATTAGCTGTGGCTA 2480
Qy 2039 TGAAGAGAAGACTGAATGGGGAAAATAGATTGGTTGGATATATGGGTCAGTCACTGAGGA 2098
Db 2481 TGAAGAGAAGACC ATGGTGCAAAAGAGATCGGTTGGATATATGGCTCGGTTACTGAGGA 2540
Qy 2099 TATCTTAAGTGGCTTCAAGATGCACTGTGAAGATGGAGATCAATTTACTGCATGCCCGT 2158
Db 2541 TATCTTAAGAGGTTCAAGATGCACT TA AGATGGAGGTCGGTCTATTGCATGCCCTT 2600
Qy 2159 AAGGGC TATC TC ATC A AAA TGTCTGATAGATTGCAGGAGGTCCT 2218
Db 2601 AAG TCAGAGAGACTCA CAAGTTCT 2660
Qy 2219 CGATG TTT TTCTC GAAATCT TCTA AGACACTGTGCTCCTGGTATGG 2278
Db 2661 CCGATGGGCTCTGG TTCCGTGGAAATATTCTGAGTAGACATTGTCCACTTTGGTATGG 2720
Qy 2279 GTTTGGAGGAGGCC TCTTAAATGGCTCCAAAGGCTTGCGTATATAAACACCATTGTGTA 2338
Db 2721 ATTTGGTGGAGGCCTGCTCAAGTGGCTTCAAAGATTGGCGTACATAAACACCATCGTTTA 2780
Qy 2339 CCCATTTACATCCGTCCCTCTCATTGGCTTATTGCACAATTCCTGCAGTTTGTCTGCTCAC 2398
Db 2781 TCCATTCACATCCGTCCCTCTTGTCGCTTATTGCACAATTCCTGCGATCTGCCTCCTTAC 2840
Qy 2399 CGGAAAATTCATCATACCAACGCTGCCAAACCTGGCAAGCATGCTGTTTCTTGGCCTCTT 2458
Db 2841 TGGAAAGTTCATCATTCCAACGCTTTCCAACTTGGCAAGCGTGCTATTTCTTGGTCTTTT 2900
Qy 2459 TATCTCCATCATTGTAACTGCGGTGCTTGAGCTAAGATGGAGCGGTGTCAGCATTGAAGA 2518
Db 2901 CCTCTCCATCATCGTCACAAGTGTGCTCGAGCTGCGATGGAGTGGCGTGAGCATTGAGGA 2960
Qy 2519 TTTATGGGTGTAATGAACAATTCTGGGTGATCGGAGGTGTTTCAGCCCATCTCTTTGCGGT 2578
Db 2961 CTGGTGGCGTAACGAGCAGTTCTGGGTGATTGGAGGTGTCTCAGCCCATCTCTTTGCCGT 3020
Qy 2579 CTTCCAGGGATTCTTAAAAATGTTGGCTGGCATCGATACGAACTTCACTGTCACAGCAAA 2638
Db 3021 CTTCCAAGGATTCCTGAAGATGTTAGCTGGCCTAGACACCAACTTCACCGTCACTACCAA 3080
Qy 2639 AGCAGCCGAAGATGCAGAATTTGGGGAGCTATATATGGTCAAGTGGACAACACTTTTGAT 2698
Db 3081 AGCAGCCGACGACGCGGAGTTTGGTGAGCTCTACATGATCAAGTGGACGACGCTGCTGAT 3140
Qy 2699 TCCTCCAACCACACTTCTCATTATCAATATGTCGGGT---TGTGCTGGATTCTCTGATGC 2755
Db 3141 ACCCCCGACCACTCTTCTCATCGTGAACATGGTGGGTGTCGTTGCCGGGTTCTCCGATGC 3200
Qy 2756 ACTCAACAAAGGATATGAAGCATGGGGGCTTCTCTTTGGCAAGGTGTTCTTTGCTTTCTG 2815
Db 3201 GCTGAACAAAGGTTATGAGGCGTGGGGACCCTCTTTGGCAAGGTCTTCTTTGCATTCTG 3260
Qy 2816 GGTGATTCTTCATCTCTATCCATTCCTTAAAGGTCTAATGGGTCGCCAAAACCTAACACC 2875
Db 3261 GGTGATTCTTCATCTCTATCCATTCCTGAAAGGTCTCATGGGTAGGCAGAACAGGACTCC 3320
Qy 2876 AACCATTGTTGTTCTCTGGTCAGTGGTGTTGGCCTCTGTCTTCTCTCTCGTTTGGGTCAA 2935
Db 3321 GACCATTGTGGTTCTCTGGTCGGTG TTCTGGCTTCTGTCTTCTCTCTCGTCTGGGTGAA 3379
Qy 2936 GATCAATCCATTCGTTAACAAAGTTGATAACACCTTGGTTGCGGAGACCTGCATTTCCAT 2995
Db 3380 GATCGATCCATTCGTGAGCAAATCCGAT---GCTGACCTCTCCCAGAGCTGCAGTTCTAT 3436

```
Qy       2996 TGATTGCTGAGCTAC 3010
              ||||| ||| | |
Db       3437 AGATTGTTGAATTGC 3451


RESULT 5
AAA67114
ID   AAA67114 standard; DNA; 3851 BP.
XX
AC   AAA67114;
XX
DT   31-OCT-2000 (first entry)
XX
DE   Pinus radiata cellulose synthase nucleotide sequence SEQ ID NO:69.
XX
KW   Eucalyptus grandis; pinus radiata; Monterey pine; modification;
KW   plant cell wall; polysaccharide; polysaccharide biosynthetic pathway;
KW   transgenic plant; ds.
XX
OS   Pinus radiata.
XX
PN   WO200022092-A2.
XX
PD   20-APR-2000.
XX
PF   08-OCT-1999;   99WO-NZ00169.
XX
PR   13-OCT-1998;   98US-0170862.
PR   11-AUG-1999;   99US-0148426.
XX
PA   (GENE-) GENESIS RES & DEV CORP LTD.
PA   (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI   Bloksberg LN;
XX
DR   WPI; 2000-339328/29.
DR   P-PSDB; AAB16307.
XX
PT   New genes encoding proteins involved in a plant polysaccharide
PT   biosynthetic pathway, useful for modulating or altering the
PT   polysaccharide content, composition or structure of the plant -
XX
PS   Claim 1; Page 71-72; 301pp; English.
XX
CC   The present invention describes isolated polynucleotides (PN) comprising
CC   a sequence selected from one of 835 nucleotide sequences given in
CC   AAA67073 to AAA67907, their (reverse) complements, sequences producing
CC   an Expectation (E) value of [illegible] computed [illegible] (BLAST) sequences,
CC   sequences at least 50% identical to these, and [illegible] 30-mers of the
CC   835 sequences or sequences that [illegible] equivalent or allelic
CC   to the 835 sequences. The polynucleotides are useful to modify the
CC   activity of a polypeptide involved in a polysaccharide biosynthetic
CC   pathway in the plant. They are especially used to modulate or alter the
CC   polysaccharide content, composition or structure of the plant. AAB16268
CC   to AAB16340 are proteins encoded by some of the polynucleotide sequence
CC   given in the present invention.
XX
SQ   Sequence 3851 BP; 1040 A; 796 C; 944 G; 1071 T; 0 other;

  Query Match           36.3%;  Score 1173;  DB 21;  Length 3851;
  Best Local Similarity 69.2%;  Pred. No. 1.5e-300;
  Matches 1692;  Conservative    0;  Mismatches  735;  Indels   18;  Gaps    6;

Qy        521 GTCTGCAGAGGCTTCGGAGCCGCTTTCAATTGTTTATCCAATTCCACGCAACAAGCTCAC 580
              | |    || ||   | |||| || ||          |||||         ||| | |
Db        941 GACCGATGAAGCAAGGCAGCCACTGTCCCGAAAAGTGCGAATTCCTTCAAGCAAAATAAA 1000

Qy        581 ACCATACAGAGCAGTGATCATTATGCGAGTGGTCATTCTGGTCCTCTTCTTCCACTTCAG 640
              |||||| ||   || |  | |  || | | | ||| |  | ||  || ||   |
Db       1001 TCCATACCGAATGGTCATTGTAATTCGACTGATAGTGCTGGTATTTTTCTCCGCTATCG 1060

Qy        641 AATAACAAATCCTGTCGATAGTGCTTTGGCCTGTGGCTTATTCTGTCACATGTGAGAT 700
                  ||||| ||  | | |  | |  ||   |     |||| || | ||  | |

Db       1061 TCTCCTGAATCCAGTGAAGAATGCATATGGGCTCTGGGCCACTTCTATCGTTTGTGAAAT 1120

Qy        701 CTGGTTTGCATTTTCTTGGGTGTTGGATCAGTTCCCCAAGTGGAATCCTGTCAATAGAGA 760
              ||||||||| || || ||| |  | |||||||| ||||||||   ||| ||| | | ||
Db       1121 CTGGTTTGCCTTGTCATGGATTCTTGATCAGTTTCCCAAGTGGTTGCCTATCAGTCGTGA 1180

Qy        761 AACGTATATCGAAAGGCTGTCGGCAAGGTATGAAAGAGAGGGTGAGCCTTCTCAGCTTGC 820
              |||||||   ||  | |||||   |||||| ||| |||| || || || ||   ||||||
Db       1181 AACGTATCCTGATCGACTGTCATTAAGGTACGAACGAGAAGGCGAACCATCAATGCTTGC 1240

Qy        821 TGGTGTGGATTTTTTCGTGAGTACTGTTGATCCGCTGAAGGAACCGCCATTGATCACTGC 880
              ||| ||  | || |||||||||| |||| |||||||| || || ||| | ||||||
Db       1241 ACCTGTTGACCTCTTTGTGAGTACTGTAGATCCACTGAAGGAGCCTCCTTTGGTTACTGC 1300

Qy        881 CAATACAGTCCTTTCCATCCTTGCTGTGGACTATCCCGTCGATAAAGTCTCCTGCTACGT 940
              |||||||||  | || |||||| | || |||| || || || || || |||||| || ||
Db       1301 CAATACAGTATTATCAATCCTTTCAGTAGACTACCCTGTAGACAATGTGTCCTGTTATGT 1360

Qy        941 GTCTGATGATGGTGCAGCTATGCTTTCATTTGAATCTCTTGTAGAAACAGCTGAGTTTGC 1000
              |||||||| || ||  | |||||| | ||||||||||     || ||  | || |||||
Db       1361 CTCTGATGACGGAGCGTCGATGCTTACTTTTGAATCTCTCTCTGAGACCTCAGAATTTGC 1420

Qy       1001 AAGGAAGTGGGTTCCGTTCTGCAAAAAATTCTCAATTGAACCAAGAGCACCGGAGTTTTA 1060
              || || ||||| || |||||||| ||||||   ||||| ||  | || || ||  | ||
Db       1421 CAGAAAATGGGTACCATTCTGCAAGAAATTCGACATTGAGCCTCGCGCTCCCGAAATCTA 1480

Qy       1061 CTTCTCACAGAAAATTGATTACTTGAAAGACAAGGTTCAACCTTCTTTCGTGAAAGAACG 1120
              ||||| |||||||||| ||  |||| |||||  |||||||  | || || ||||| ||
Db       1481 TTTCTCTCAGAAAATTGACTATCTGAAGGACAAATTTCAACCCACCTTTGTCAAAGAGCG 1540

Qy       1121 TAGAGCAATGAAAAGGGATTATGAAGAGTACAAAGTCCGAGTTAATGCCCTGGTAGCAAA 1180
              | || ||||| || || |||||||| | ||| || ||  | |||    |||| |||||
Db       1541 CCGTGCCATGAAGAGAGAATATGAAGAATTCAAGGTGCGCATCAATCGGTTGGTTGCAAA 1600

Qy       1181 GGCTCAGAAAACACCTGAAGAAGGATGGACTATGCAAGATGGAACACCTTGGCCTGGGAA 1240
              |||    |||   ||  | |||||||||| |||||||| || || ||| | ||  |    | |  ||||
Db       1601 GGCCTCTAAAGTGCCCAAGGAAGGATGGACAATGCAAGACGGTACGCCTTGGCCTGGTAA 1660

Qy       1241 TAACACACGTGATCACCCTGGGCATGATTCAGGTCTTCCTTGGGAAATACTGGGAGCTCG 1300
              ||| || ||||| || ||| || ||||| || || ||| | ||  |    | |  ||||
Db       1661 TAATACCCGTGACCATCCT-GGTATGATCCAAGTGTTCTTGGGTCACAGTGGCGGCCTCG 1719

Qy       1301 TGACATTGAAGGAAATGAACTACCTCGTCTAGTATATGTCTCCAGGGAGAAGAGACCTGG 1360
              |||  ||||| ||||| || ||||| |||||||||| ||  | ||||||||||||||
Db       1720 ATACA--GAAGGCAATGAGCTTCCTCGGCTAGTATATGTTTCTCGTGAGAAGAGACCTGG 1777

Qy       1361 CTACCAGCACCACAAAAAGGCTGGTGCAGAAAATGCTCTGGTGAGAGTGTCTGCAGTACT 1420
              | ||||' |' || ||||| |||||    ||||| |||| | || ||||| || ||
Db       1778 TTTCCAGCATCACAAGAAGGCCGGTGCCATGAATGCTTTGGTTCGGGTTTCTGCTGTGCT 1837

Qy       1421 CACAAATGCTCCCTACATCCTCAATGTTGATTGTGATCACTATGTAAACAATAGCAAGGC 1480
              || |||||||| |  || || ||| | ||||||||||||||  | ||||||||||||||
Db       1838 CACCAATGCTCCATTTATGCTGAATCTGGATTGTGATCACTACATTAACAATAGCAAGGC 1897

Qy       1481 TGTTCGAGAGGCAATGTGCATCCTGATGGACCCACAAGTAGGTCGAGATGTATGCTATGT 1540
              |  | || |  |||||  | ||||||| || || || ||  || | || || |||||
Db       1898 AATCAGGGAAGGCATGTGCTTTATGATGGATCCTCAGGTTGGGAGAAAAGTCTGTTATGT 1957

Qy       1541 GCAGTTCCCTCAGAGGTTTGATGGCATAGATAAGAGTGATCGCTACGCCAATCGTAACGT 1600
              || |||||||||| || ||||| || |||    | ||| || ||||||||||| |||
Db       1958 CCAATTCCCTCAGAGATTCGATGGTATTGATCGCAATGACCGTTACGCCAATCGAAACAC 2017

Qy       1601 AGTTTTCTTTGATGTTAACATGAAAGGGTTGGATGGCATTCAAGGACCAGTATACGTAGG 1660
              || |||||||| | |||||||||  |||||| |||||||| || ||||| || ||
Db       2018 CGTATTCTTTGATATCAACATGAAAGGTCTGGATGGAATTCAAGGGCCTGTATATGTGGG 2077

Qy       1661 AACTGGTTGTGTTTTCAACAGGCAAGCACTTTACGGCTACGGGCCTCCTTCTATGCCCAG 1720
              |||||| ||  | |||| || ||||| || || || || ||||||||       || |
Db       2078 AACTGGATGCATGTTCAGAAGACAAGCTCTATATGGGTATGGGCCTCCCAAAGGCCCAAA 2137

Qy       1721 CTTACGCAAGAGAAAGGATTCTTCATCCTGCTTCTCATGTTGCTGCCCCCTCAAAGAAGAA 1780
                 | ||||   | ||  | |    ||  || | |||||| | ||    ||||||
Db       2138 ACGTCCCAAG---ATGGTGACCTGTGATTGTCTCCCTTGTTGCGGTCCTCGTAAGAAGTC 2194
```

```
1781 GCCTGCTCAAGATCCAGCTGAGGTATACAGAGATGCAAAAAGAGAGGATCTCAATGCTGC 1840
2195 TCCGAAGAAAAAATAGTAGCAAGAAAAGTGCAGGAATCCCAGCTCCCGCCTACAATCTGGA 2254

1841 CATATTTAATCTTACAGAGATTGATAATTATGACGAGCATGAAAGGTCAATGCTGATCTC 1900
2255 CGGGATCGAGGAAGGAGTAGAAGGTTATGATGAC------GAAAGAGCATTGTTGATGAG 2308

1901 CCAGTTGAGCTTTGAGAAAACTTTTGGCTTATCTTCTGTCTTCATTGAGTCTACACTAAT 1960
2309 CCAACTAGACTTCGAGAAGAAGTTTGGCCAGTCTTCAGCTTTTGTTCAATCCACTCTGAT 2368

1961 GGAGAATGGAGGAGTACCCGAGTCTGCCAACTCACCACCATTCATCAAGGAAGCGATTCA 2020
2369 GGAGAATGGTGGTGTTCCGCAAACAGCAAATCCAGCTGAATTGTTGAAGGAGGCTATTCA 2428

2021 AGTCATCGGCTGTGGCTATGAAGAGAAGACTGAATGGGGAAAACAGATTGGTTGGATATA 2080
2429 TGTCATCAGCTGTGGATATGAAGACAAAACGGAATGGGGAAAAGAGCTTGGATGGATCTA 2488

2081 TGGGTCAGTCACTGAGGATATCTTAAGTGGCTTCAAGATGCACTGCCGAGGATGGAGATC 2140
2489 TGGATCAGTCACAGAGGACATTCTGACTGGATTCAAGATGCACACTCGAGGCTGGCGGTC 2548

2141 AATTTACTGCATGCCCGTAAGGCCTGCATTCAAAGGATCTGCACCCATCAACCTGTCTGA 2200
2549 CATTTACTGTATGCCCAAACGAGCAGCATTCAAAGGGTCTGCTCCAATCAATCTATCAGA 2608

2201 TAGATTGCACCAGGTCCTCCGATGGGCTCTTGGTTCTGTGGAAATTTTCTTTAGCAGACA 2260
2609 CCGTTTGAACCAGGTGTTGCGTTGGGCTTTGGGATCAGTAGAAATTTTCATGAGCAGACA 2668

2261 CTGTCCCCTCTGGTACGGGTTTGGAGGAGGCCGTCTTAAATGGCTCCAAAGGCTTGCGTA 2320
2669 TTGCCCAATCTGGTATGGCTATGGGGGAG---GTCTGAAATGGCTTGAAAGATTTGCCTA 2725

2321 TATAAACACCATTGTGTACCCATTTACATCCCTCCCTCTCATTGCCTATTGCACAATTCC 2380
2726 TATCAACACCATTGTCTATCCATTCACCTCTCTTCCACTCATTGCCTATTGCACACTTCC 2785

2381 TGCAGTTTGTCTGCTCACCGGAAAATTCATCATACCAACGCTCTCAAACCTGGCAAGCAT 2440
2786 AGCCGTCAGTTTGCTCACTGGCAAATTTGTGATCCCTCAGATCAGTACTTTTGCAAGTCT 2845

2441 GCTGTTTCTTGGCCTCTTTATCTCCATCATTGTAACTGCGGTGCTTGAGCTAAGATGGAG 2500
2846 ATTTTTAATAGCTCTTTTCATCTCAATTTTTGCCACTGGTATTCTGAAAATGAGGTGGAG 2905

2501 CGGTGTCAGCATTGAAGATTTATGGCGTAATGAACAATTCTGGGTGATCGGAGGTGTTTC 2560
2906 TGGAGTGAGCATTGAAGAATGGTGGCGAAATGAACAGTTCTGGGTTATTGGAGGGGTTTC 2965

2561 AGCCCATCTCTTTGCGGTCTTCCAGGGATTCTTAAAAATGTTGGCTGGCATCGATACGAA 2620
2966 TGCACATTTTTTTGCAGTTATTCAAGGTCTGCTCAAGGTACTGGCAGGCATTGATACAAA 3025

2621 CTTCACTGTCACAGCAAAAGCAGCCGAAGATGCAGAATTTGGGGAGCTATATATGGTCAA 2680
3026 TTTCACAGTCACTGCCAAGGCATCAGATGACGGTGAGTTTGGGGAACTGTATGCATTCAA 3085

2681 GTGGACAACACTTTTGATTCCTCCAACCACACTTCTCATTATCAA---TATGTCGGGTTG 2737
3086 ATGGACCACACTCCTCATTCCTCCTACAACCCTGCTTGTCATCAACCTTGTGGGGGTGGT 3145

2738 TGCTGGATTCTCTGATGCACTCAACAAAGGATATGAAGCATGGGGGCCTCTCTTTGGCAA 2797
3146 TGTTGGCGTAGCAGATGCAATCAACAATGGATTTCAGTCATGGGGTCCTCTCTTGGGTAA 3205

2798 GGTGTTCTTTGCTTTCTGGGTGATTCTTCATCTCTATCCATTCCTTAAAGGTCTAATGGG 2857
3206 GCTTTTCTTTGCATTCTGGGTCATTGTGCACCTGTATCCTTTCCTCAAGGGTCTCATGGG 3265
```

```
Qy       2858 TCGCCAAAACCTAACACCAACCATTGTTGTTCTCTGGTCAGTGCTGTTGGCCTCTGTCTT 2917
Db       3266 CAGGCAGAACCGAACACCCACCATCGTGGTTATTTGGTCAATTCTGCTGGCATCTGTTTT 3325

Qy       2918 CTCTCTCGTTTGGGTCAAGATCAATCCATTCGTTAACAAAGTTGA 2962
Db       3326 CTCTCTTTTCTGGGTAAGAATTGATCCTTTCTTGAGTAAGGTTAA 3370
```

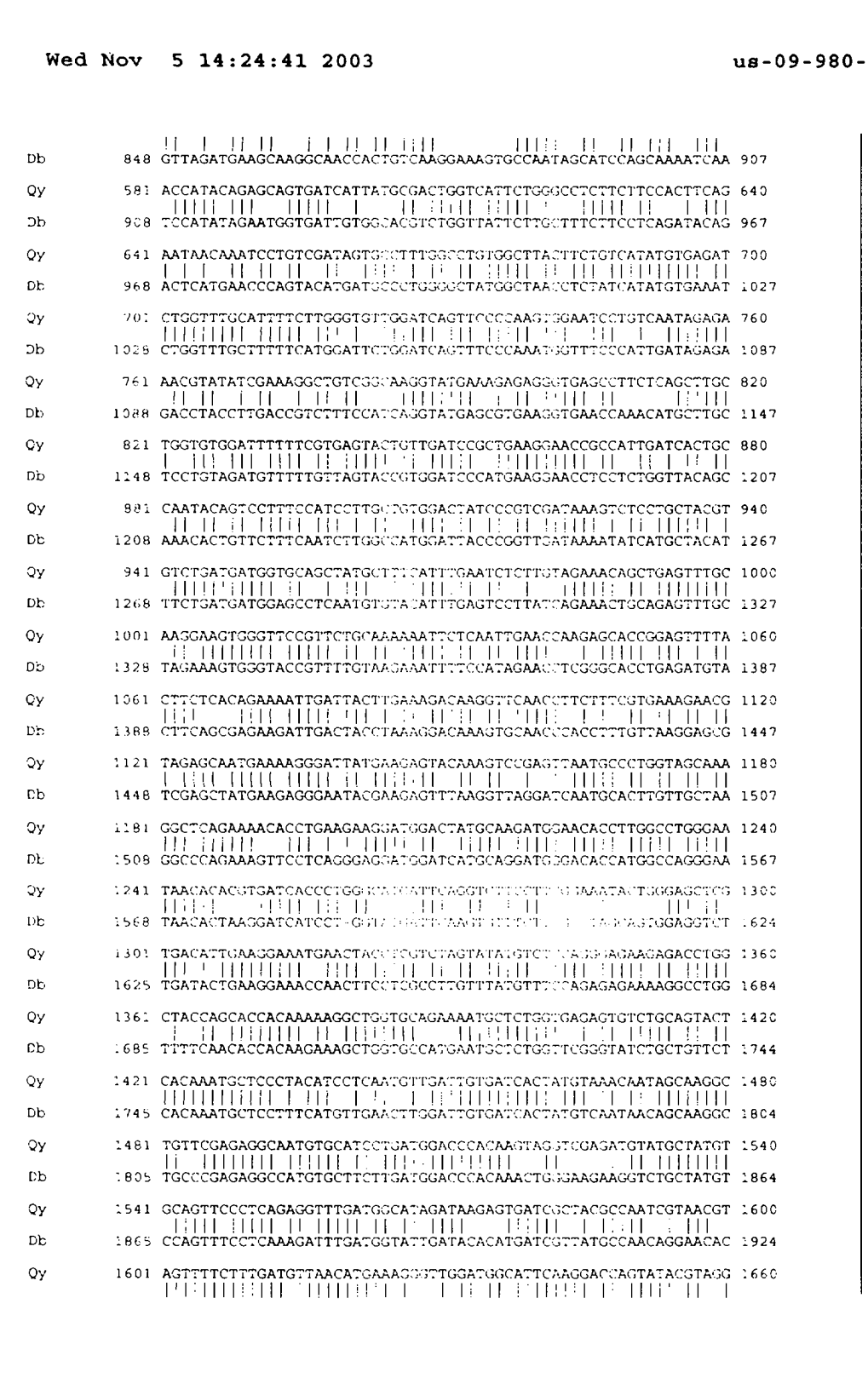
RESULT 6
AAZ58268
ID AAZ58268 standard; cDNA; 3517 BP.
XX
AC AAZ58268;
XX
DT 08-MAY 2000 (first entry)
XX
DE Soybean cellulose synthase cDNA clone sdp2c.pk005.o22.
XX
KW Soybean; cellulose synthase; CS; transferase; plant; ss.
XX
OS Glycine max.
XX

FH	Key	Location/Qualifiers
FT	CDS	158..3278
FT		/*tag= a
FT		/transl_except= (pos:758..761, aa:Xaa)
FT		/note= "apparent 1 base addition, alters the reading frame"

XX
PN WO200004166-A2.
XX
PD 27-JAN-2000.
XX
PF 13-JUL 1999; 99WO-US15871.
XX
PR 14-JUL-1998; 98US-009.844
XX
PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX
PI Allen SM, Fader GM, Falco SC, Kinney AJ, Lightner JE, Miao G;
PI Rafalski JA, Thorpe CJ;
XX
DR WPI; 2000-182491/16.
DR P-PSDB; AAY58837.
XX
PT Nucleic acid fragments encoding cellulose biosynthetic enzyme useful as
PT probes for isolating cDNAs and genes encoding homologous proteins, for
PT producing transgenic plants
XX
PS Claim 3; Page 55-56; 93pp; English.
XX
CC The present sequence is that of full length cDNA clone
CC sdp2c.pk005.o22 encoding soybean cellulose synthase (CS, see AAY58837).
CC The cDNA clone was isolated from a cDNA library produced from
CC soybean pod tissue. The invention relates to isolated nucleic acid
CC fragments encoding plant CS and to CS polypeptides. It also relates
CC to the construction of a chimeric gene encoding all or a portion of
CC the CS, in sense or antisense orientation, where expression of the
CC gene results in altered levels of the CS in transformed host cells.
CC The host cells can be used to screen compounds for their ability to
CC inhibit CS activity. CS nucleic acids are also useful for producing
CC transgenic plants having altered levels of CS, and hence altered
CC levels of fibre. CS may also serve as a target for the development
CC of novel herbicides.
XX
SQ Sequence 3517 BP; 978 A; 686 C; 841 G; 1012 T; 0 other;

Query Match 35.7%; Score 1152.6; DB 21; Length 3517;
Best Local Similarity 69.1%; Pred. No. 3.6e-295;
Matches 1686; Conservative 0; Mismatches 699; Indels 54; Gaps 6;

```
Qy        521 GTCTGCAGAGGCTTCGGAGCCGCTTTCAATTGTTTATCCAATTCCACGCAACAAGCTCAC 580
```


Db 848 GTTAGATGAAGCAAGGCAACCACTGTCAAGGAAAGTGCCAATAGCATCCAGCAAAATCAA 907
Qy 581 ACCATACAGAGCAGTGATCATTATGCGACTGGTCATTCTGGGCCTCTTCTTCCACTTCAG 640
Db 908 TCCATATAGAATGGTGATTGTGGCACGTCTGGTTATTCTTGCTTTCTTCCTCAGATACAG 967
Qy 641 AATAACAAATCCTGTCGATAGTGCCTTTGGCCTGTGGCTTACTTCTGTCATATGTGAGAT 700
Db 968 ACTCATGAACCCAGTACATGATGCCCTGGGCCTATGGCTAACCTCTATCATATGTGAAAT 1027
Qy 701 CTGGTTTGCATTTTCTTGGGTGTTGGATCAGTTCCCCAAGTGGAATCCTGTCAATAGAGA 760
Db 1028 CTGGTTTGCTTTTTCATGGATTCTGGATCAGTTTCCCAAATGGTTTCCCATTGATAGAGA 1087
Qy 761 AACGTATATCGAAAGGCTGTCGGAAGGTATGAAAGAGAGGGTGAGCCTTCTCAGCTTGC 820
Db 1088 GACCTACCTTGACCGTCTTTCCATCAGTATGAGCGTGAAGGTGAACCAAACATGCTTGC 1147
Qy 821 TGGTGTGGATTTTTTCGTGAGTACTGTTGATCCGCTGAAGGAACCGCCATTGATCACTGC 880
Db 1148 TCCTGTAGATGTTTTTGTTAGTACCGTGGATCCCATGAAGGAACCTCCTCTGGTTACAGC 1207
Qy 881 CAATACAGTCCTTTCCATCCTTGCTGTGGACTATCCCGTCGATAAAGTCTCCTGCTACGT 940
Db 1208 AAACACTGTTCTTTCAATCTTGGCCATGGATTACCCGGTTGATAAAATATCATGCTACAT 1267
Qy 941 GTCTGATGATGGTGCAGCTATGCTTTCATTTGAATCTCTTGTAGAAACAGCTGAGTTTGC 1000
Db 1268 TTCTGATGATGGAGCCTCAATGTGTATATTTGAGTCCTTATCAGAAACTGCAGAGTTTGC 1327
Qy 1001 AAGGAAGTGGGTTCCGTTCTGCAAAAAATTCTCAATTGAACCAAGAGCACCGGAGTTTTA 1060
Db 1328 TAGAAAGTGGGTACCGTTTTGTAAGAAATTTTCCATAGAACCTCGGGCACCTGAGATGTA 1387
Qy 1061 CTTCTCACAGAAAATTGATTACTTGAAAGACAAGGTTCAACCTTCTTTCGTGAAAGAACG 1120
Db 1388 CTTCAGCGAGAAGATTGACTACCTAAAGGACAAAGTGCAACCCACCTTTGTTAAGGAGCG 1447
Qy 1121 TAGAGCAATGAAAAGGGATTATGAAGAGTACAAAGTCCGAGTTAATGCCCTGGTAGCAAA 1180
Db 1448 TCGAGCTATGAAGAGGGAATACGAAGAGTTTAAGGTTAGGATCAATGCACTTGTTGCTAA 1507
Qy 1181 GGCTCAGAAAACACCTGAAGAAGGATGGACTATGCAAGATGGAACACCTTGGCCTGGGAA 1240
Db 1508 GGCCCAGAAAGTTCCTCAGGGAGGATGGATCATGCAGGATGGGACACCATGGCCAGGGAA 1567
Qy 1241 TAACACACGTGATCACCCTGG[illegible]AAATACTGGGAGCTCG 1300
Db 1568 TAACACTAAGGATCATCCT-G[illegible]AAGTGGAGGTCT 1624
Qy 1301 TGACATTGAAGGAAATGAACTAC[illegible]AGAAGAGACCTGG 1360
Db 1625 TGATACTGAAGGAAACCAACTTCCTGGCCTTGTTTATGTTTCCAGAGAGAAAAGGCCTGG 1684
Qy 1361 CTACCAGCACCACAAAAAGGCTGGTGCAGAAAATGCTCTGGTGAGAGTGTCTGCAGTACT 1420
Db 1685 TTTTCAACACCACAAGAAAGCTGGTGCCATGAATGCCTGGTTCGGGTATCTGCTGTTCT 1744
Qy 1421 CACAAATGCTCCCTACATCCTCAATGTTGATTGTGATCACTATGTAAACAATAGCAAGGC 1480
Db 1745 CACAAATGCTCCTTTCATGTTGAACTTGGATTGTGATCACTATGTCAATAACAGCAAGGC 1804
Qy 1481 TGTTCGAGAGGCAATGTGCATCCTGATGGACCCACAAGTAGTCGAGATGTATGCTATGT 1540
Db 1805 TGCCCGAGAGGCCATGTGCTTCTTGATGGACCCACAAACTGGGAAGAAGGTCTGCTATGT 1864
Qy 1541 GCAGTTCCCTCAGAGGTTTGATGGCATAGATAAGAGTGATCGCTACGCCAATCGTAACGT 1600
Db 1865 CCAGTTTCCTCAAAGATTTGATGGTATTGATACACATGATCGTTATGCCAACAGGAACAC 1924
Qy 1601 AGTTTTCTTTGATGTTAACATGAAAGGGTTGGATGGCATTCAAGGACCAGTATACGTAGG 1660
Db 1925 AGTTTTCTTTGATATTAACATGAAGGGTCTAGATGGTATTCAAGGTCCTGTATATGTGGG 1984
Qy 1661 AACTGGTTGTGTTTTCAACAGGCAAGCACTTTACGGCTACGGGCCTCCTTCTATGCCCAG 1720
Db 1985 GACTGGATGTGTTTTCAGGAGGCAAGCTTTGTATGGCTATAATCCTCCCAAGGGTCCAAA 2044
Qy 1721 CTTACGCAAGAGAAAGGATTCTTCATCCTGCTTCTCATGTTGCTGCCCCTCAAAGAAGAA 1780
Db 2045 GCGTCCAAAAATGGTAAGCTGTGATTGTTGCCCGTGTTTTGGAAGCCGCAAGAAGTA--- 2101
Qy 1781 GCCTGCTCAAGATCCAGCTGAGGTATACAGAGATGCAAAAAGAGAGGATCTCAATGCTGC 1840
Db 2102 ----------TAAGGAGAAGAATGATGCAAATGGAGAGGCT---------- 2132
Qy 1841 CATATTTAATCTTACAGAGATTGATAATTATGACGAGCATGAAAGGTCAATGCTGATCTC 1900
Db 2133 ----GCAAGCCTAAAAGGGATGGATGATGACAAAGAG----------GTGTTGATGTC 2176
Qy 1901 CCAGTTGAGCTTTGAGAAAACTTTTGGCTTATCTTCTGTCTTCATTGAGTCTACACTAAT 1960
Db 2177 CCAAATGAATTTTGAGAAGAAATTTGGACAATCCTCTATTTTTGTGACTTCTACCTTGAT 2236
Qy 1961 GGAGAATGGAGGAGTACCCGAGTCTGCCAACTCACCACCATTCATCAAGGAAGCGATTCA 2020
Db 2237 GGAAGAGGGTGGTGTGCCTCCTTCTTCAAGTCCAGCTGCCCTGCTTAAAGAAGCCATTCA 2296
Qy 2021 AGTCATCGGCTGTGGCTATGAAGAGAAGACTGAATGGGGAAAACAGATTGGTTGGATATA 2080
Db 2297 TGTGATTAGCTGTGGATATGAAGATAAAACTGAATGGGGACTTGAGCTTGGTTGGATCTA 2356
Qy 2081 TGGGTCAGTCACTGAGGATATCTTAAGTGGCTTCAAGATGCACTGCCGAGGATGGAGATC 2140
Db 2357 TGGATCTATCACAGAAGATATTCTAACAGGTTTTAAGATGCATTGCCGTGGGTGGAGGTC 2416
Qy 2141 AATTTACTGCATGCCCGTAAGGCCTGCATTCAAAGGATCTGCACCCATCAACCTGTCTGA 2200
Db 2417 CATTTATTGTATGCCAAAGAGAGCTGCATTCAAGGGTACTGCTCCTATCAACTTGTCAGA 2476
Qy 2201 TAGATTGCACCAGGTCCTCCGATGGGCTCTTGGTTCTGTGGAAATTTTCTTTAGCAGACA 2260
Db 2477 TCGTCTCAACCAGGTTCTTCGTTGGGCACTTGGTTCCATTGAGATTTTCTTTAGTCACCA 2536
Qy 2261 CTGTCCCCTCTGGTACGGGTTTGGAGGAGGCCGTCTTAAATGGCTCCAAAGGCTTGCGTA 2320
Db 2537 TTGCCCTCTATGGTATGGCTTCAAGGAAAAGAAGCTAAAGTGGCTTGAGAGATTTGCCTA 2596
Qy 2321 TATAAACACCATTGTGTACCCATTTACATCCCTCCCTCTCATTGCCTATTGCACAATTCC 2380
Db 2597 TGCAAACACAACTGTCTATCCATTCACCTCCATTCCTCTAGTTGCCTACTGTATTCTTCC 2656
Qy 2381 TGCAGTTTGTCTGCTCACCGGAAAATTCATCATACCAACGCTCTCAAACCTGGCAAGCAT 2440
Db 2657 AGCAGTTTGTTTACTCACTGACAAATTCATCATGCCACCGATTAGCACCTTTGCTGGTTT 2716
Qy 2441 GCTGTTTCTTGGCCTCTTTATCTCCATCATTGTAACTGCGGTGCTTGAGCTAAGATGGAG 2500
Db 2717 GTACTTTGTTGCTCTCTTCTCCTCAATCATTGCAACTGGTATTCTTGAGTTGAAATGGAG 2776
Qy 2501 CGGTGTCAGCATTGAAGATTTATGGCGTAATGAACAATTCTGGGTGATCGGAGGTGTTTC 2560
Db 2777 TGGAGTGAGCATTGAGGAATGGTGGAGAAATGAGCAGTTTTGGGTCATTGGTGGTGTATC 2836
Qy 2561 AGCCCATCTCTTTGCGGTCTTCCAGGGATTCTTAAAAATGTTGGCTGGCATCGATACGAA 2620
Db 2837 AGCTCACCTCTTTGCTGTTATACAAGGTCTGCTAAAGGTTCTGGCTGGAATTGACACCAA 2896
Qy 2621 CTTCACTGTCACAGCAAAAGCAGCCGAAGATGCAGAATTTGGGGAGCTATATATGGTCAA 2680
Db 2897 TTTCACTGTTACATCAAAGGCAACAGATGATGAAGAGTTTGGAGAATTGTACACCTTTAA 2956
Qy 2681 GTGGACAACACTTTTGATTCCTCCAACCACACTT---CTCATTATCAATATGTCGGGTTG 2737
Db 2957 GTGGACTACACTCTTGATTCCTCCAACCACTATTTTGATCATTAACATTGTTGGTGTTGT 3016

```
      2738 TGCTGGATTCTCTGATGCACTCAACAAAGGATATGAAGCATGGGGGCCTCTCTTTGGCAA 2797
           ||||||| |||| ||||| | ||||| || ||  || | ||||| || |||||||| ||
      3017 TGCTGGAATCTCAGATGCCATAAACAATGGGTACCAATCCTGGGGACCACTCTTTGGAAA 3076

      2798 GGTGTTCTTTGCTTTCTGGGTGATTCTTCATCTCTATCCATTCCTTAAAGGTCTAATGGG 2857
           | | |||||| | |||||||||||| | |||||||||||||||||||||||| | |||||
      3077 GCTCTTCTTTTCCTTCTGGGTGATTGTCCATCTCTATCCATTCCTTAAAGGTTTGATGGG 3136

      2858 TCGCCAAAACCTAACACCAACCATTGTTGTTCTCTGGTCAGTGCTGTTGGCCTCTGTCTT 2917
           ||||||||| |  ||||| |||||||||||  | ||||||||||| |||||||| | ||
      3137 TCGCCAAAATCGCACACCCACCATTGTTGTGATTTGGTCAGTGCTATTGGCCTCTATTTT 3196

      2918 CTCTCTCGTTTGGGTCAAGATCAATCCATTCGTTAACAA 2956
           |||  |  ||||||| |  ||  ||||||| ||   |||
      3197 CTCCTTACTTTGGGTAAGAATTGATCCATTTGTCCTCAA 3235


ESULT 7
AZ99515
D   AAZ99515 standard; DNA; 3753 BP.
X
    AAZ99515;
X
    03-JUL-2000  (first entry)

    DNA encoding a maize cellulose synthase.

K
N   Maize; cellulose synthase; stalk quality; stand; silage; cellulose;
N   transgenic plant; plant breeding marker; ss.
X
S   Zea mays.
X
H   Key             Location/Qualifiers
T   CDS             184..3405
T                   /*tag=  a
T                   /product= "cellulose synthase"
T                   /transl_except= (pos: 916..918, aa: Xaa)
T                   /note= "no termination codon given; Xaa is an
T                          unspecified amino acid"
X
X
N   WO200009706-A2.
X
D   24-FEB-2000.
X
F   16-AUG-1999;   99WO-US18760.
X
R   17-AUG-1998;   98US-0096822.
X
A   (PION-) PIONEER HI-BRED INT INC.
X
I   Dhugga KS,  Helentjaris TG,  Bowen BA,  Wang X;
X
R   WPI; 2000-224343/19.
R   P-PSDB; AAY84115.
X
T   New genes which encode maize cellulose synthase polypeptides in plants
T   useful for modulating the expression of cellulose synthase in plants
T   and to produce transgenic plants expressing the novel protein  -
X
S   Claim 1; Page 144-149; 119pp; English.
X
C   The present sequence encodes a maize cellulose synthase polypeptide.
C   The cellulose synthase can be used for the improvement of stalk quality
C   for improved stand or silage. It also provides an increased concentration
C   of cellulose in the pericarp, hardening the kernel and improving its
C   handling ability. The sequences are used to produce transgenic plants
C   and seeds expressing the cellulose synthase. The polynucleotide is
C   used for modulating, preferably increasing, the level of the synthase
C   in a plant cell. The plants are preferably monocots. The polynucleotide
C   is also used as a probe or primer in the detection quantitation or
C   isolation of gene transcripts. The probes are useful in detecting
C   deficiencies in the level of mRNA in screenings for desired transgenic
C   plant, for detecting mutations in the gene, for monitoring upregulation
C   of expression or changes in enzyme activity in screening assays of
C   compounds, for detection of any number of allelic variants of the gene,
C   or for use as molecular markers in plant breeding programs. The
C   isolated nucleic acids of the present invention can also be used for
C   recombinant expression of their encoded polypeptides or for use as
C   immunogens in the preparation and/or screening of antibodies. The
C   proteins can be employed in assays for enzyme agonists or antagonists
C   of enzyme function or for use of immunogens or antigens to obtain
C   antibodies specifically immunoreactive with a protein.
XX
SQ  Sequence 3753 BP; 844 A; 805 C; 565 G; 1048 T; 1 other;

  Query Match             35.5%;  Score 1148.8;  DB 21;  Length 3753;
  Best Local Similarity   67.1%;  Prec. No. 3.8e 294;
  Matches 1734;  Conservative    0;  Mismatches  837;  Indels   15;  Gaps    7;

Qy        374 TGAAATGAATGATGAATATGGAAGTTAATTTTGAAGAATCGGGTGAAGAGCTGTAAGGA 433
              ||  |  |  |  |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db        740 TGGTCTTAATA-T-TAATGTAT-AAATATCTGAGAGCTGGAGGGTTAAACAGGACAA 839

Qy        434 TAAAGAGAATAAAAAAGAAAAAGAGAATTCTTAAGCTGAAACTGAACCAGTTCAAGTTCC 493
              ||  ||  |  |  | | | | |  |||  |  | |  |  |  |  |  |  |  |
Db        840 AAATATGATGCAAGTGACTAATAAATATCAAGGCTAGAGGAGGAGACATGGAGGGGAC 899

Qy        494 TACAGAACAGCAGATGGAAGAGAAACCGTCTGCAGAGGCTTCGGAGCCGCTTTCAATTGT 553
              |       |       |  | ||  |  || ||  |  |  |  |  |  |  |
Db        900 TGGCTCAAATGGAGAANATATGCAAATGGTTTATGATGCACGGCTACCTTTGAGCCGTAT 959

Qy        554 TTATCCAATTCCAGCGCAACAAGCTCACACCATACAGAGCAGTGATCATTATGCGACTGGT 613
              ||  ||  |  ||| ||  ||  |  ||  | | ||||||| | |  ||  |
Db        960 CGTGCCAATTTCCCTCAAACCAGCTCAACCTTTACCGGGTAGTGATCATTCTCCGTCTTAT 1019

Qy        614 CATTCTGGGCCTCTTCTTCCACTTCAGAATAACAAATCCTGTCGATAGTGCCTTTGGCCT 673
              ||| ||| ||  ||| | |  |  |  ||| ||  |  |  |  |  |  |  |  |
Db       1020 CATCCTGTGCTTCTTCTTCCAGTATTGTGTCAGTCATCCAGTGCGTGATGCTTATGGATT 1079

Qy        674 GTGGCTTACTTCTGTCATATGTGAGATCTGGTTTGCATTTTCTTGGGTGTTGGATCAGTT 733
              |||||  |||| || ||  | | ||||  ||| |  |||||  |  |  |||  |||
Db       1080 ATGGCTAGTATCTGTTATCTGCGAGGTCTGGTTTGCCTTGTCTTGGCTTCTAGATCAGTT 1139

Qy        734 CCCCAAGTGGAATCCTGTCAATAGAGAAACGTATATCGAAAGGCTGTCGGCAAGGTATGA 793
              ||| || ||| |  ||  |||  |  ||  |  ||  | ||  ||||  |  |  |||||  |
Db       1140 CCCAAAATGGTATCCAATCAACCGTGAGACATATCTTGACAGGCTTGCATTGAGGTATGA 1199

Qy        794 AAGAGAGGGTGAGCCTTCTCAGCTTGCTGGTGTGGATTTTTTCGTGAGTACTGTTGATCC 853
              |  ||||  |||  | ||||  |||     |  ||| |  ||||  |||| |  ||||
Db       1200 TAGAGAGGGAGAGCCATCACAGCTGGCTCCCATTGATGTCTTCGTCAGTACAGTGGATCC 1259

Qy        854 GCTGAAGGAACCGCCATTGATCACTGCCAATACAGTCCTTTCCATCCTTGCTGTGGACTA 913
              |  ||||  |||  |||||||  ||||  |  ||  |  ||||  |||  |  |||||  ||
Db       1260 ATTGAAGGAACCTCCACTGATCACAGCCAACACTGTTTTGTCCATTCTTTCTGTGGATTA 1319

Qy        914 TCCCGTCGATAAAGTCTCCTGCTACGTGTCTGATGATGGTGCAGCTATGCTTTCATTTGA 973
              || |  || |||||  ||  ||  ||  ||| ||||||  ||||||||||  | |||||
Db       1320 CCCTGTTGACAAAGTGTCATGCTATGTTTCTGATGATGGTTCAGCTATGCTGACTTTTGA 1379

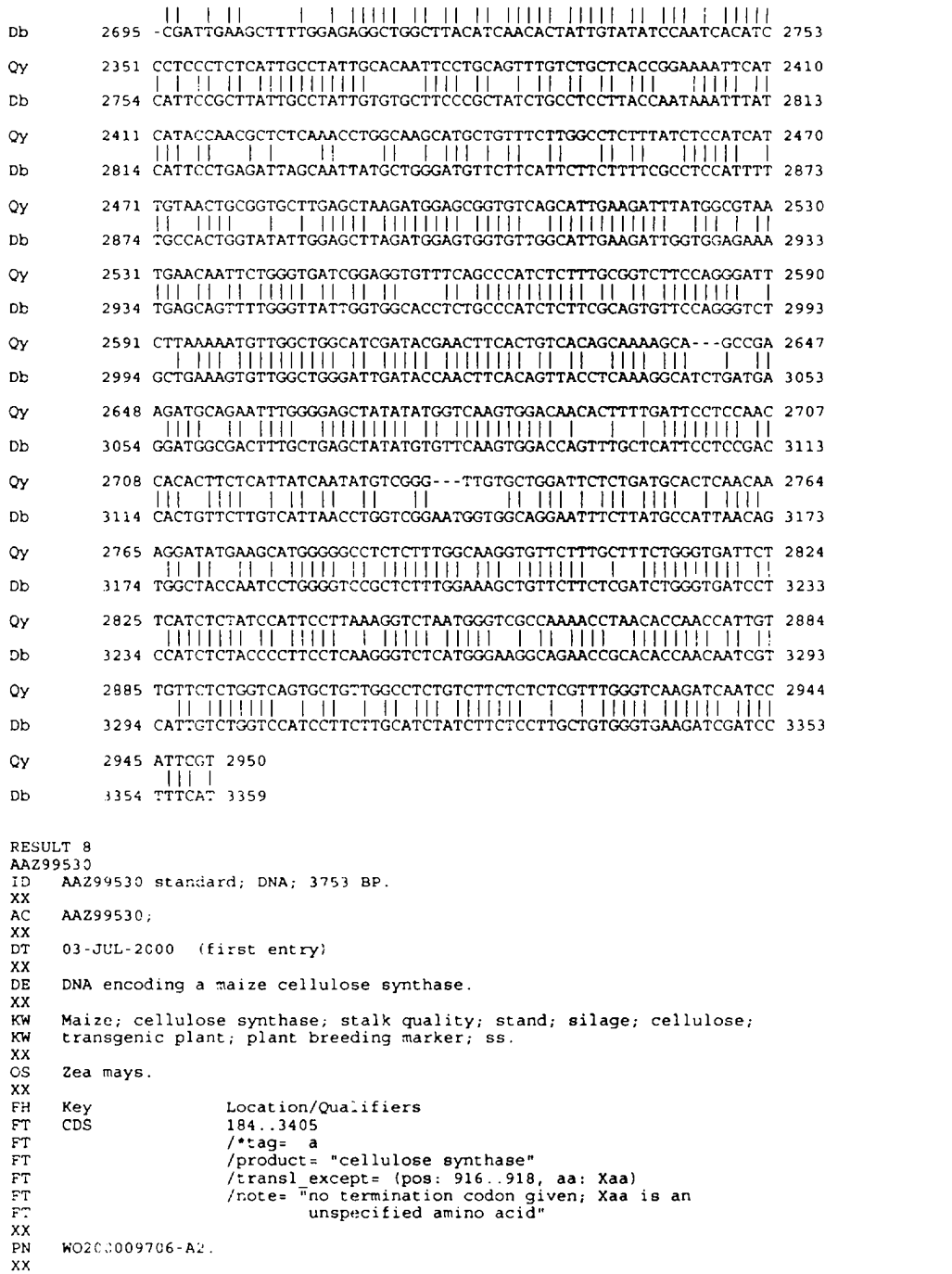
Qy        974 ATCTCTTGTAGAAACAGCTGAGTTTGCAAGGAAGTGGGTTCCGTTCTGCAAAAAATTCTC 1033
              |  |||   |||  |  ||  ||  ||  |||  |||||||  ||  |  ||  |  |
Db       1380 GTCTCTCTCAGAAACCGCAGAATTTGCTAGAAAGTGGGTTCCCTTTTGTAAGAAGCACAA 1439

Qy       1034 AATTGAACCAAGAGCACCGGAGTTTTACTTCTCACAGAAAATTGATTACTTGAAAGACAA 1093
              |||||||||||||  ||  || ||||||||  |  || |||||  ||||  |||  ||||
Db       1440 TATTGAACCAAGAGCTCCAGAATTTTACTTTGCTCAAAAAATAGATTACCTGAAGGACAA 1499

Qy       1094 GGTTCAACGTTCTTTCGTGAAAGAACGTAGAGCAATGAAAAGGGATTATGAAGAGTACAA 1153
              |  ||||||||  |  ||  ||  |||  |  |  ||||||| |||||  ||||||||  | |||
Db       1500 AATTCAACCTTCATTTGTTAAGGAAAGACGCGCAATGAAGAGGGAGTATGAAGAATTCAA 1559

Qy       1154 AGTCCGAGTTAATGCCCTGGTAGCAAAGGCTCAGAAAACACCTGAAGAAGGATGGACTAT 1213
              |||  ||  |  ||||||||  ||  ||  |  |  ||||  |  ||||||||  ||  ||||  ||
Db       1560 AGTAAGAATCAATGCCCTTGTTGCCAAAGCACAGAAAGTGCCTGAAGAGGGGTGGACCAT 1619
```





RESULT 8
AAZ99530
ID AAZ99530 standard; DNA; 3753 BP.
XX
AC AAZ99530;
XX
DT 03-JUL-2000 (first entry)
XX
DE DNA encoding a maize cellulose synthase.
XX
KW Maize; cellulose synthase; stalk quality; stand; silage; cellulose;
KW transgenic plant; plant breeding marker; ss.
XX
OS Zea mays.
XX
FH Key Location/Qualifiers
FT CDS 184..3405
FT /*tag= a
FT /product= "cellulose synthase"
FT /transl_except= (pos: 916..918, aa: Xaa)
FT /note= "no termination codon given; Xaa is an
FT unspecified amino acid"
XX
PN WO2003009706-A2.
XX

D 24-FEB-2000.

F 16-AUG-1999; 99WO-US18760.

R 17-AUG-1998; 98US-0096822.

A (PION-) PIONEER HI-BRED INT INC.

I Dhugga KS, Helentjaris TG, Bowen BA, Wang X;

R WPI; 2000-224343/19.
R P-PSDB; AAY84120.

T New genes which encode maize cellulose synthase polypeptides in plants
T useful for modulating the expression of cellulose synthase in plants
T and to produce transgenic plants expressing the novel protein -

S Claim 1; Page 184-188; 119pp; English.

C The present sequence encodes a maize cellulose synthase polypeptide.
C The cellulose synthase can be used for the improvement of stalk quality
C for improved stand or silage. It also provides an increased concentration
C of cellulose in the pericarp, hardening the kernel and improving its
C handling ability. The sequences are used to produce transgenic plants
C and seeds expressing the cellulose synthase. The polynucleotide is
C used for modulating, preferably increasing, the level of the synthase
C in a plant cell. The plants are preferably monocots. The polynucleotide
C is also used as a probe or primer in the detection quantitation or
C isolation of gene transcripts. The probes are useful in detecting
C deficiencies in the level of mRNA in screenings for desired transgenic
C plant, for detecting mutations in the gene, for monitoring upregulation
C of expression or changes in enzyme activity in screening assays of
C compounds, for detection of any number of allelic variants of the gene,
C or for use as molecular markers in plant breeding programs. The
C isolated nucleic acids of the present invention can also be used for
C recombinant expression of their encoded polypeptides or for use as
C immunogens in the preparation and/or screening of antibodies. The
C proteins can be employed in assays for enzyme agonists or antagonists
C of enzyme function or for use of immunogens or antigens to obtain
C antibodies specifically immunoreactive with a protein.

X

Q Sequence 3753 BP; 944 A; 805 C; 965 G; 1038 T; 1 other;

Query Match 35.5%; Score 1148.8; DB 21; Length 3753;
Best Local Similarity 67.1%; Pred. No. 3.8e-294;
Matches 1734; Conservative 0; Mismatches 837; Indels 15; Gaps 7;

```
Qy        374 TGAAATGAATGATGAATATGGGAATCCAATTTGGAAGAATCGGGTGAAGAGCTGTAAGGA 433
              ||   | |||  ||   |  ||||   ||      |||   ||  |   |      |  |
Db        780 TGGGCTTAATAGTGTTGACTGGAAGGAAAGAGTTGAGAGCTGGAGGGTTAAACAGGACAA 839

Qy        434 TAAAGAGAACAAGAAGAAAAAGAGAAGTCCTAAGGCTGAAACTGAACCAGCTCAAGTTCC 493
              ||   ||   |   ||  || | |  |||  |||||  |   | |  |       | |   |
Db        840 AAATATGATGCAAGTGACTAATAAATATCCAGAGGCTAGAGGAGGAGACATGGAGGGGAC 899

Qy        494 TACAGAACAGCAGATGGAAGAGAAACCGTCTGCAGAGGCTTCGGAGCCGCTTTCAATTGT 553
              |      | |          |   | ||  |  ||  ||  ||    |   ||  |       | |
Db        900 TGGCTCAAATGGAGAAANATATGCAAATGGTTGATGATGCACGGCTACCTTTGAGCCGTAT 959

Qy        554 TTATCCAATTCCACGCAACAAGCTCACACCATACAGAGCAGTGATCATTATGCGACTGGT 613
                  |||||| |    |||  ||||||  |  |||  | |  |||||||||| | ||  ||  |
Db        960 CGTGCCAATTTCCTCAAACCAGCTCAACCTTTACCGGGTAGTGATCATTCTCCGTCTTAT 1019

Qy        614 CATTCTGGGCCTCTTCTTCCACTTCAGAATAACAAATCCTGTCGATAGTGCCTTTGGCCT 673
              |||  |||  ||  |||||||||| |   |  | |   |||| ||   |  |||  |  |||  |
Db        1020 CATCCTGTGCTTCTTCTTCCAGTATCGTGTCAGTCATCCAGTGCGTGATGCTTATGGATT 1079

Qy        674 GTGGCTTACTTCTGTCATATGTGAGATCTGGTTTGCATTTTCTTGGGTGTTGGATCAGTT 733
              |||||    |||||  ||  ||  ||| ||||||||||  ||  ||||||  |  | ||||||||
Db        1080 ATGGCTAGTATCTGTTATCTGCGAGGTCTGGTTTGCCTTGTCTTGGCTTCTAGATCAGTT 1139

Qy        734 CCCCAAGTGGAATCCTGTCAATAGAGAAACGTATATCGAAAGGCTGTCGGCAAGGTATGA 793
              |||  ||  |||  ||||  ||||  |  ||  ||  |||  |  ||  |||||  |    ||||||||
```

```
Db        1140 CCCAAAATGGTATCCAATCAACCGTGAGA ATATCTTGACAGGCTTGCATTGAGGTATGA 1199

Qy        794 AAGAGAGGGTGAGCCTTCTCAGCTTGCTG NTGGATTTTTTCGTGAGTACTGTTGATCC 853
              |||||||| |.||| || ||||| ||     ||| | ||||| |||| || ||||
Db        1200 TAGAGAGGGAGAGCCATCACAGCTGGCTC CATTGATGTCTTCGTCAGTACAGTGGATCC 1259

Qy        854 GCTGAAGGAACCGCCATTGATCACTGCCAATACAGTCCTTTCCATCCTTGCTGTGGACTA 913
              |||||:||| .|| .|||| |||. || ||  | ||||| ||, |||||||  |
Db        1260 ATTGAAGGAACCTCCACTGATCACAGCCAACACTGTTTTGTCCATTCTTTCTGTGGATTA 1319

Qy        914 TCCCGTCGATAAAGTCTCCTGCTACGTGT TGATGATGGTGCAGCTATGCTTTCATTTGA 973
              || .. || |||||  | |  |  ||,| |||||||  ||||||||||,  | ||||
Db        1320 CCCTGTTGACAAAGTGTCATGCTATGTTT TGATGATGGTTCAGCTATGCTGACTTTTGA 1379

Qy        974 ATCTCTTGTAGAAACAGCTGAGTTTGCAA GAAGTGGGTTCCGTTCTGCAAAAAATTCTC 1033
              |.|||   ||||:| ||  ||   ||  :|||||||||  || || || ||    |
Db        1380 GTCTCTCTCAGAAACCGCAGAATTTGCTA GAAAGTGGGTTCCCTTTTGTAAGAAGCACAA 1439

Qy        1034 AATTGAACCAAGAA TACCGGAGTTTTACT CTCACAGAAAATTGATTACTTGAAAGACAA 1093
               | |||  |||  | |  ||  :||     |  |||| |||||  ||  ||||
Db        1440 TATTGAA AAGA TTC AGAATTTTAC  C   AAAAAATAGATTACTTGAAGGACAA 1499

Qy        1094 GGTTCAACTTTCTTTCGTGAAAGAACGTA A CAATGAAAAGGGATTATGAAGAGTACAA 1153
               ||||.|||||, || || |: ||| :  | |  |||| |||||  |||||:|| | ||
Db        1500 AATTCAACCTTCATTTGTTAAGGAAAGA CGCGCAATGAAGAGGGAGTATGAAGAATTCAA 1559

Qy        1154 AGTCCGAGTTAATGCCCTGGTAGCAAAG CTCAGAAAACACCTGAAGAAGGATGGACTAT 1213
               ||| || | |||  ||| || |  :| |.  |||||   |||||||| || |||||  ||
Db        1560 AGTAAGAATCAATGCCCTTGTTGCCAAA CACAGAAAGTGCCTGAAGAGGGGTGGACCAT 1619

Qy        1214 GCAAGATGGAACACCTTGGCCTGGGAATAACACACGTGATCACCCTGGGCATGATTCAGG 1273
               |  ||||||  | ||||||||||,|| |  | || || ||| |||||||||||
Db        1620 GGCTGATGGAACTGCATGGCCTGGGAATAATGCTAGGGACCATCCT-GGCATGATTCAGG 1678

Qy        1274 TCTTCCTTGGGAAATACTGGGAGCTCGTGACATTGAAGGAAATGAACTACCTCGTCTAGT 1333
               | ||| | |||  |  | | |||  |:| ||| ||||||||  ||| |||||  ||
Db        1679 TTTTCTTGGGGCACAGTGGTGGGCTC--GACACTGATGGAAATGAGTTACCACGTCTTGT 1736

Qy        1334 ATATGTCTCCAGGGAGAAGAGACCTGGCTACCAGCACCACAAAAAGGCTGGTGCAGAAAA 1393
               |||||||  | || |||||||| |:||  |:|| |||| |||||||||||  ||
Db        1737 CTATGTCTCTCGTGAAAAGAGACCAGGCTTTCAGCATCACAAGAAGGCTGGTGCAATGAA 1796

Qy        1394 TGCTCTGGTGAGAGTGTCTGCAGTACTCACAAATGCTCCCTACATCCTCAATGTTGATTG 1453
               ||| |||  |  | |||| || ||:|||  | |||| |||||||||||  |||||
Db        1797 TGCGCTGATTCGTGTATCTGCTGTGCTGACAAATGGTGCCTATCTTCTCAATGTGGATTG 1856

Qy        1454 TGATCACTATGTAAACAATAGCAAGGCTGTTCGAGAGGCAATGTGCATCCTGATGGACCC 1513
               || || ||  | || |  |||| || |. |||| |||||||||  || ||||||  ||
Db        1857 CGACCATTACTTCAATAGCAGCAAAGCTCTTAGAGAAGCAATGTGCTTCATGATGGATCC 1916

Qy        1514 ACAAGTAGGTCGAGATGTATGCTATGTGCAGTTCCCTCAGAGGTTTGATGGCATAGATAA 1573
                  ||||  :  |  ||,|||| | | || |||||  |||||||||  ||
Db        1917 GGCTCTAGGAAGGAAAACTTGTTATGTACAATTTCCACAGAGATTTGATGGCATTGACTT 1976

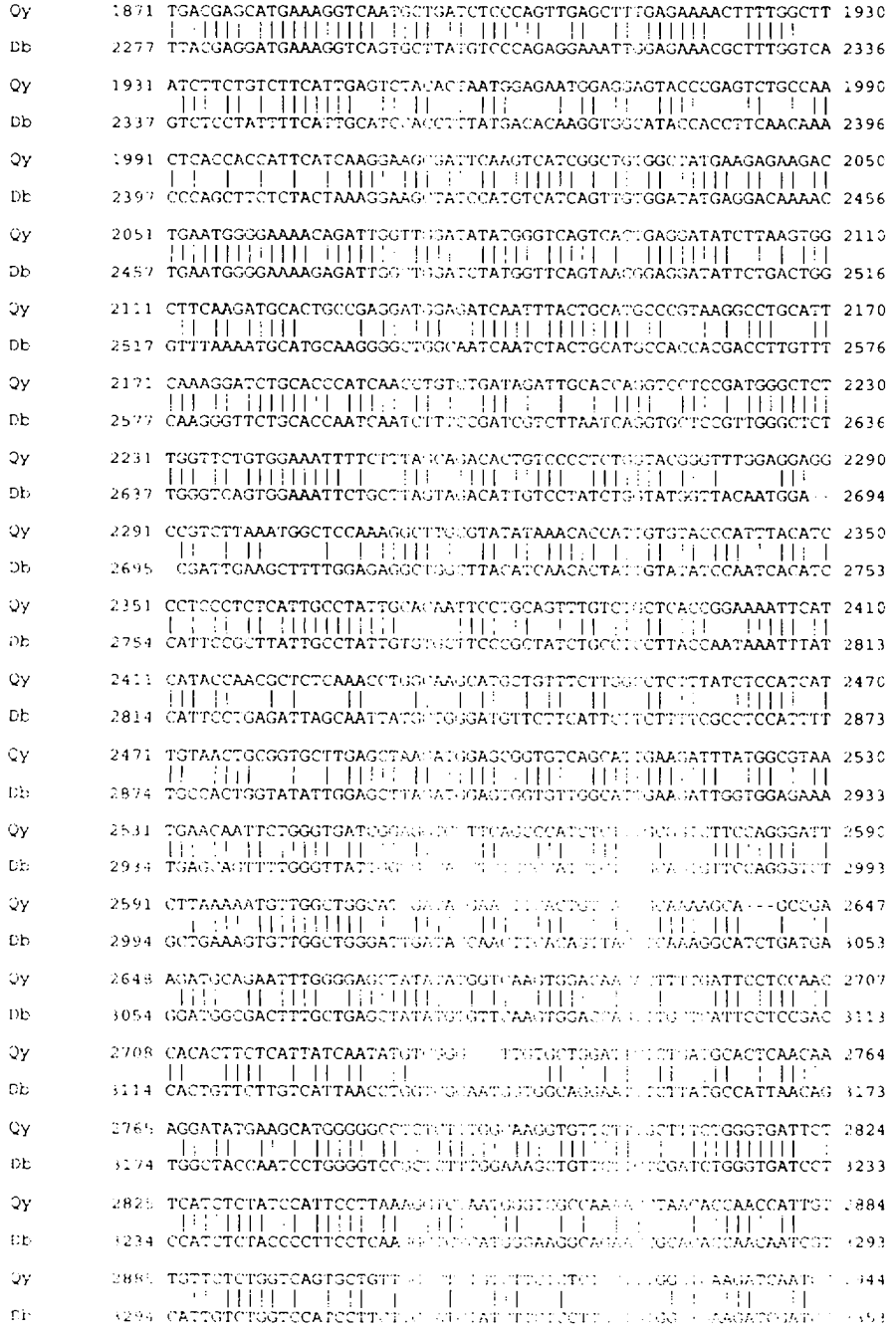
Qy        1574 GAGTGATCGCTACGCCAATCGTAACGTAGTTTTCTTTGATGTTAACATGAAAGGGTTGGA 1633
               |  ||,|| ||  | |||| :|| ||. ||||  |||| | |||||||||  ||||
Db        1977 GCACGATCGATATGCTAATCGGAACATATGTTTTCTTTGATATCAACATGAAAGGTCTGGA 2036

Qy        1634 TGGCATTCAAGGACCAGTATACGTAGGAACTTGTTGTGTTTTCAACAGGCAAGCACTTTA 1693
               |||||||| || |:|| ||||| .|||  |  ||   |||||| || || ||  | ||
Db        2037 TGGCATTCAGGGTCCAGTTTACGTGGGAACAGGATGCTGTTTCAATAGACAGGCTTTGTA 2096

Qy        1694 CGGCTACGGGCCTCCTTCTATGCCCAGCTTACGCAAGAGAAAGGATTCTTCATCCTGCTT 1753
               || |||  ,||  |  |                 || |||  |  |  |||
Db        2097 TGGATACGATCCTGTTTTGACTGAAGCTGATCTGGAGCCAAACATTGTTATTAAGAGCTG 2156

Qy        1754 CTCATGTTGCTGCCCCTCAAAGAAGAAGCCTGCTCAAGAT-CCAGCTGAGGTATACAGAG 1812
               ||  || |  |  |||| : ||| |  '  |||  |  |||| !|
Db        2157 CTGTGGTAGAAGGAAGAAAAAGAACAAGAGTTATATGGATAGTCAAAGCCGTATTATGAA 2216

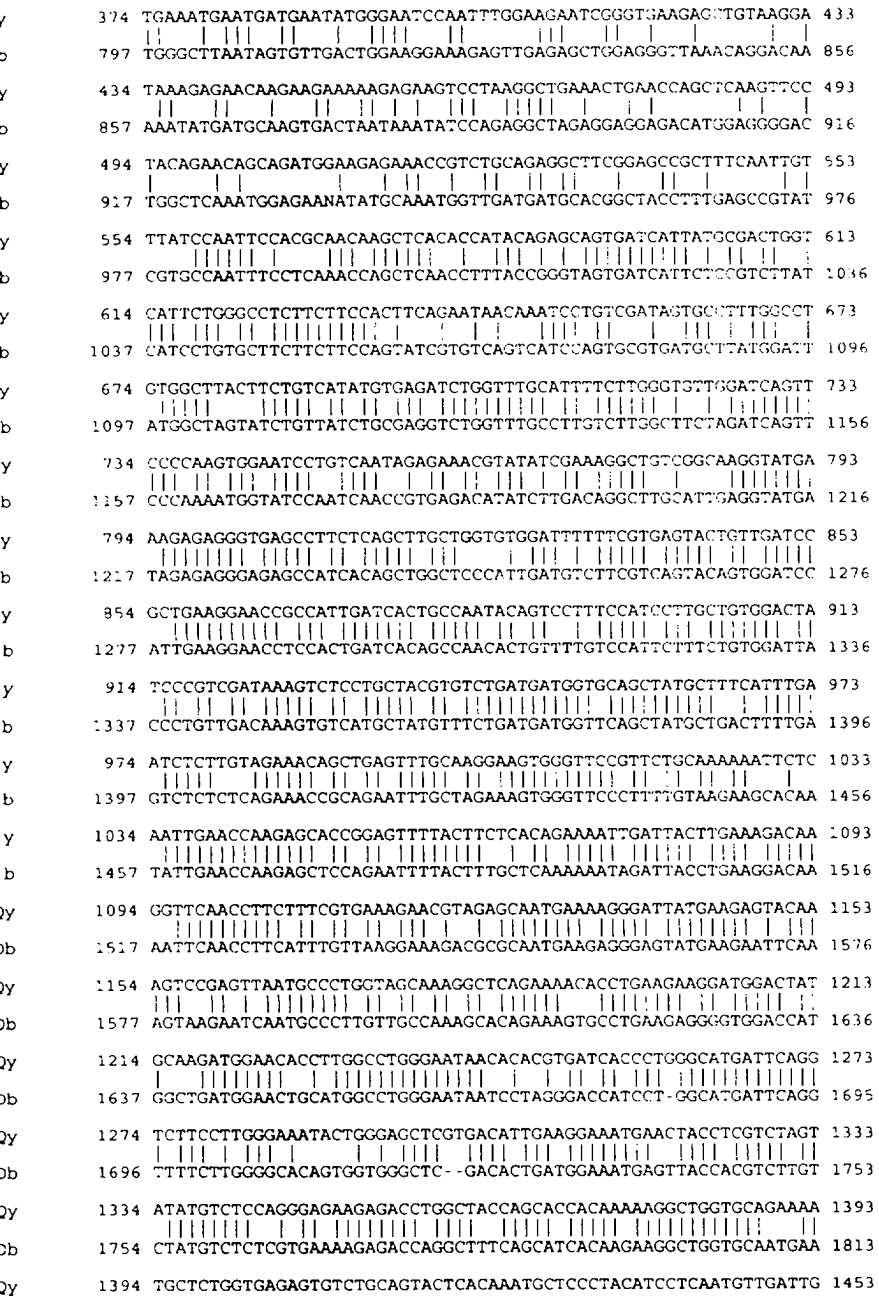
Qy        1813 ATGCAAAAAAGAGAGGATCTCAATGCTGCCATATTTA--ATCTTACAGAGATTGATAATTA 1870
               | | |  |  ||| || |||  |  |  |  |||| |  | |
Db        2217 GAGAACAGAATCTTCAGCTCCCATCTTCAATATGGAAGACATCGAAGAGGGTATTGAAGG 2276
```

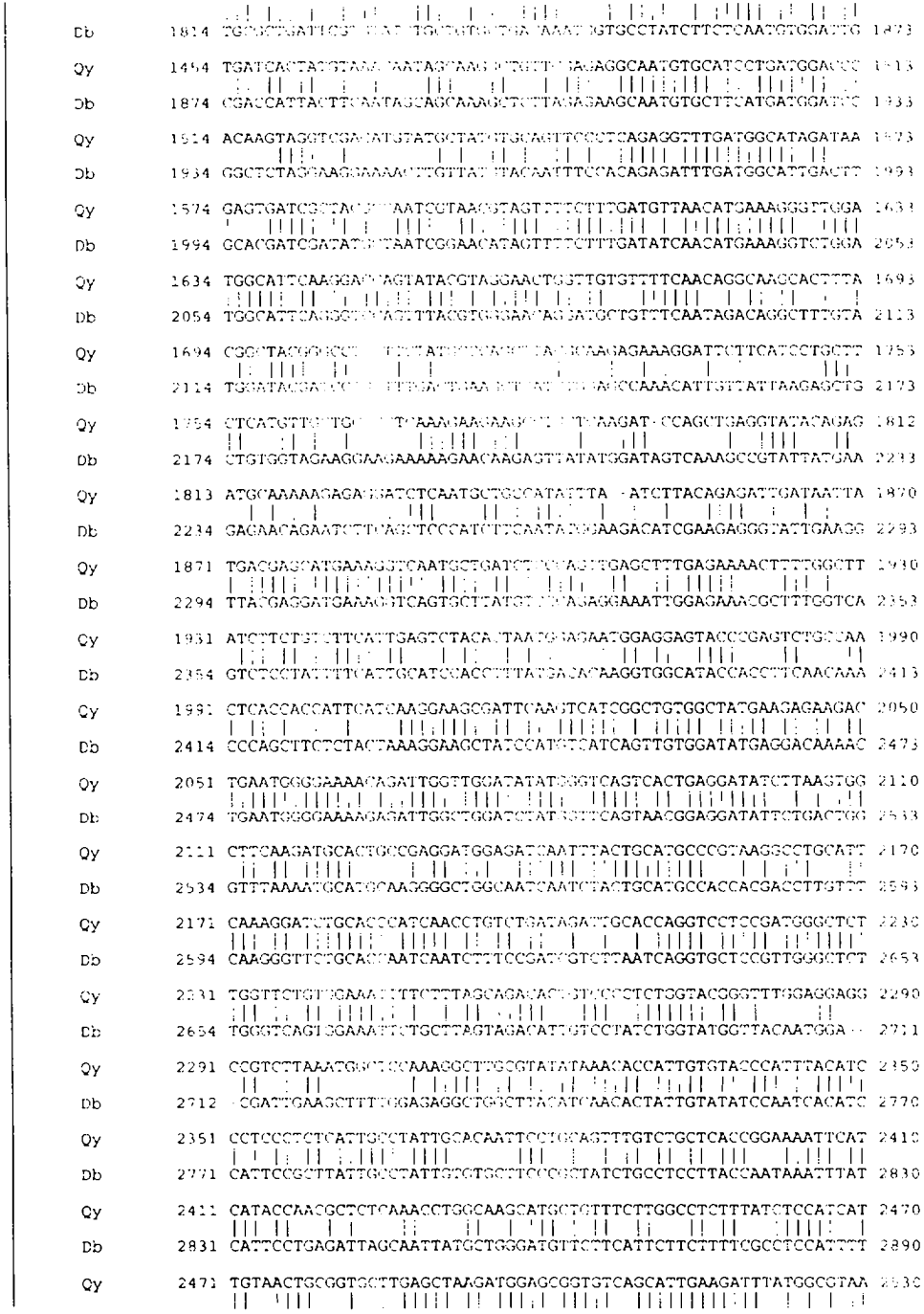


```
Qy       2945 ATTCGT 2950
                 ||| |
Db       3354 TTTCAT 3359


RESULT 9
AAZ99497
ID   AAZ99497 standard; DNA; 3780 BP.
XX
AC   AAZ99497;
XX
DT   03-JUL-2000  (first entry)
XX
DE   DNA encoding a maize cellulose synthase.
XX
KW   Maize; cellulose synthase; stalk quality; stand; silage; cellulose;
KW   transgenic plant; plant breeding marker; ss.
XX
OS   Zea mays.
XX
FH   Key             Location/Qualifiers
FT   CDS             201..3422
FT                   /*tag= a
FT                   /product= "cellulose synthase"
FT                   /transl_except= (pos: 933..935, aa: Xaa)
FT                   /note= "no termination codon given; Xaa is an
FT                         unspecified amino acid"
XX
PN   WO200009706-A2.
XX
PD   24-FEB-2000.
XX
PF   16-AUG-1999;   99WO US18760.
XX
PR   17-AUG-1998;   98US-0096822.
XX
PA   (PION-) PIONEER HI-BRED INT INC.
XX
PI   Dhugga KS,  Helentjaris TG,  Bowen BA,  Wang X;
XX
DR   WPI; 2000-224343/19.
DR   P-PSDB; AAY84109.
XX
PT   New genes which encode maize cellulose synthase polypeptides in plants
PT   useful for modulating the expression of cellulose synthase in plants
PT   and to produce transgenic plants expressing the novel protein -
XX
PS   Claim 1; Page 97-102; 119pp; English.
XX
CC   The present sequence encodes a maize cellulose synthase polypeptide.
CC   The cellulose synthase can be used for the improvement of stalk quality
CC   for improved stand or silage. It also provides an increased concentration
CC   of cellulose in the pericarp, hardening the kernel and improving its
CC   handling ability. The sequences are used to produce transgenic plants
CC   and seeds expressing the cellulose synthase. The polynucleotide is
CC   used for modulating, preferably increasing, the level of the synthase
CC   in a plant cell. The plants are preferably monocots. The polynucleotide
CC   is also used as a probe or primer in the detection quantitation or
CC   isolation of gene transcripts. The probes are useful in detecting
CC   deficiencies in the level of mRNA in screenings for desired transgenic
CC   plant, for detecting mutations in the gene, for monitoring upregulation
CC   of expression or changes in enzyme activity in screening assays of
CC   compounds, for detection of any number of allelic variants of the gene,
CC   or for use as molecular markers in plant breeding programs. The
CC   isolated nucleic acids of the present invention can also be used for
CC   recombinant expression of their encoded polypeptides or for use as
CC   immunogens in the preparation and/or screening of antibodies. The
CC   proteins can be employed in assays for enzyme agonists or antagonists
CC   of enzyme function or for use of immunogens or antigens to obtain
CC   antibodies specifically immunoreactive with a protein.
XX
SQ   Sequence 3780 BP; 944 A; 817 C; 976 G; 1040 T; 3 other;
```

Query Match 35.5%; Score 1148.8; DB 21; Length 3780;
Best Local Similarity 67.1%; Pred. No. 3.8e-294;
Matches 1734; Conservative 0; Mismatches 837; Indels 15; Gaps 7;

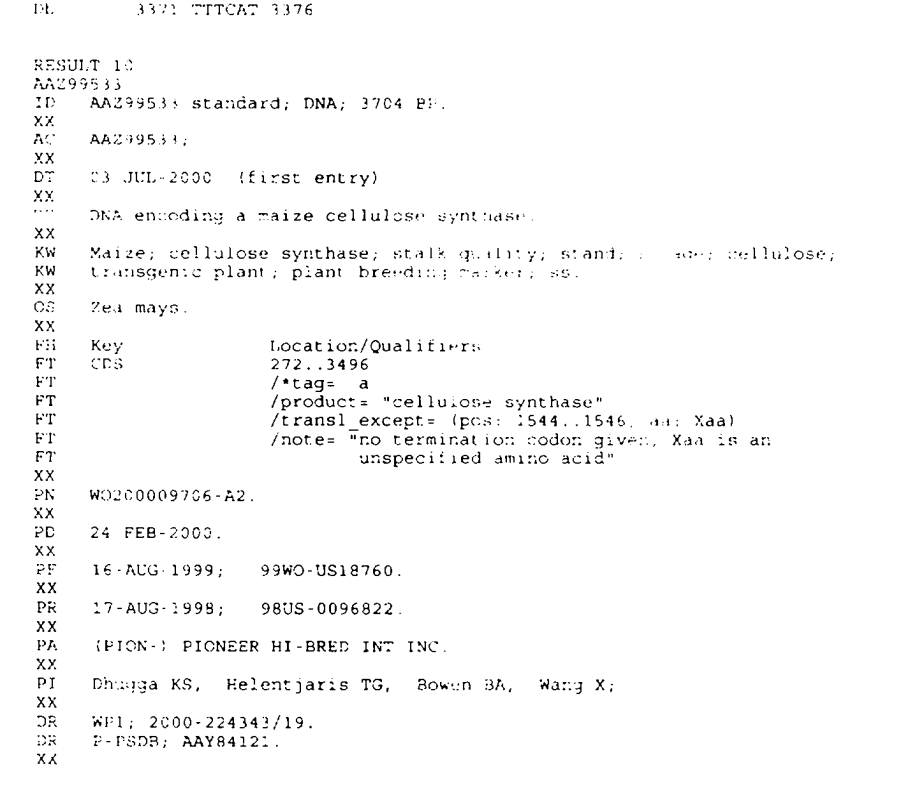


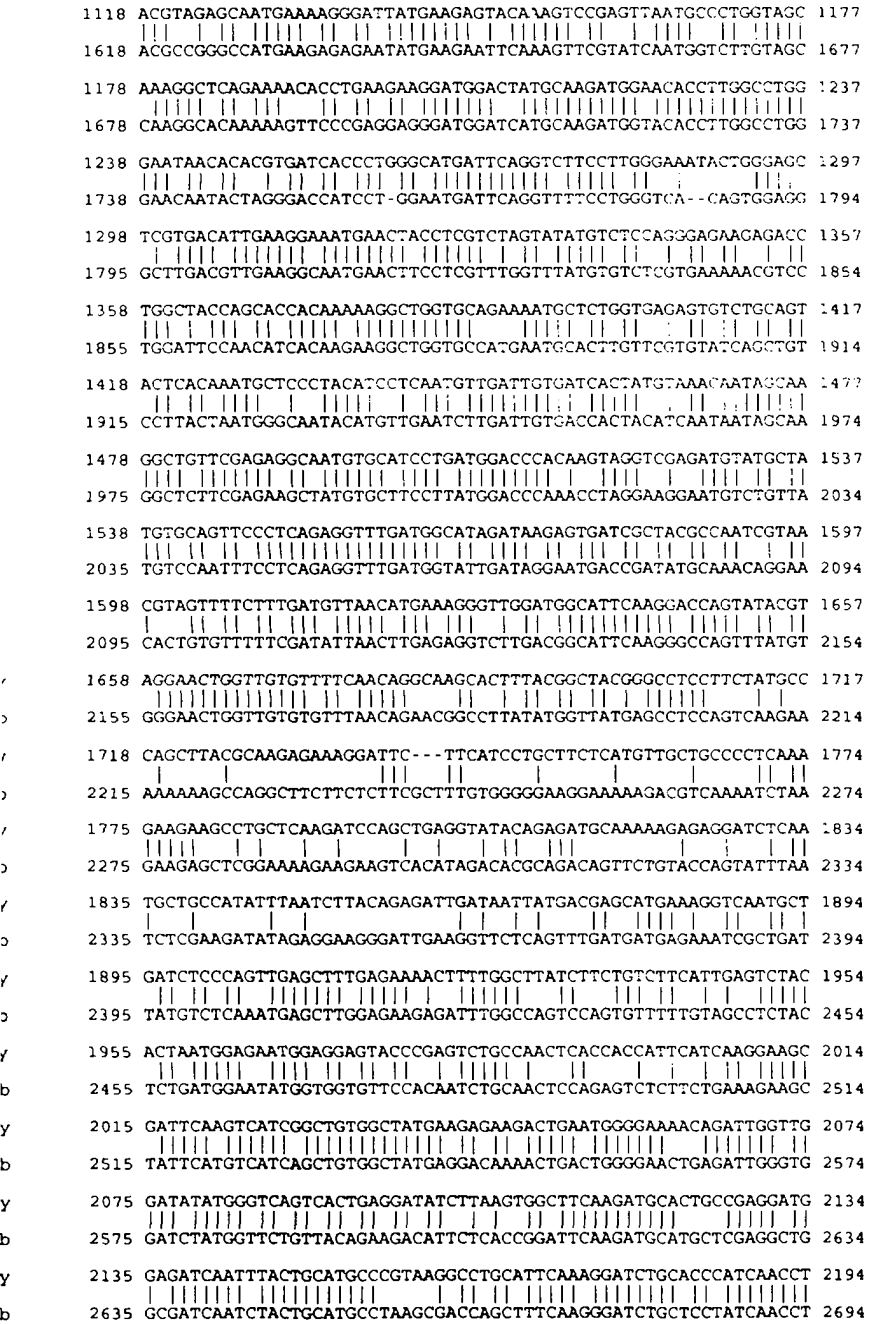


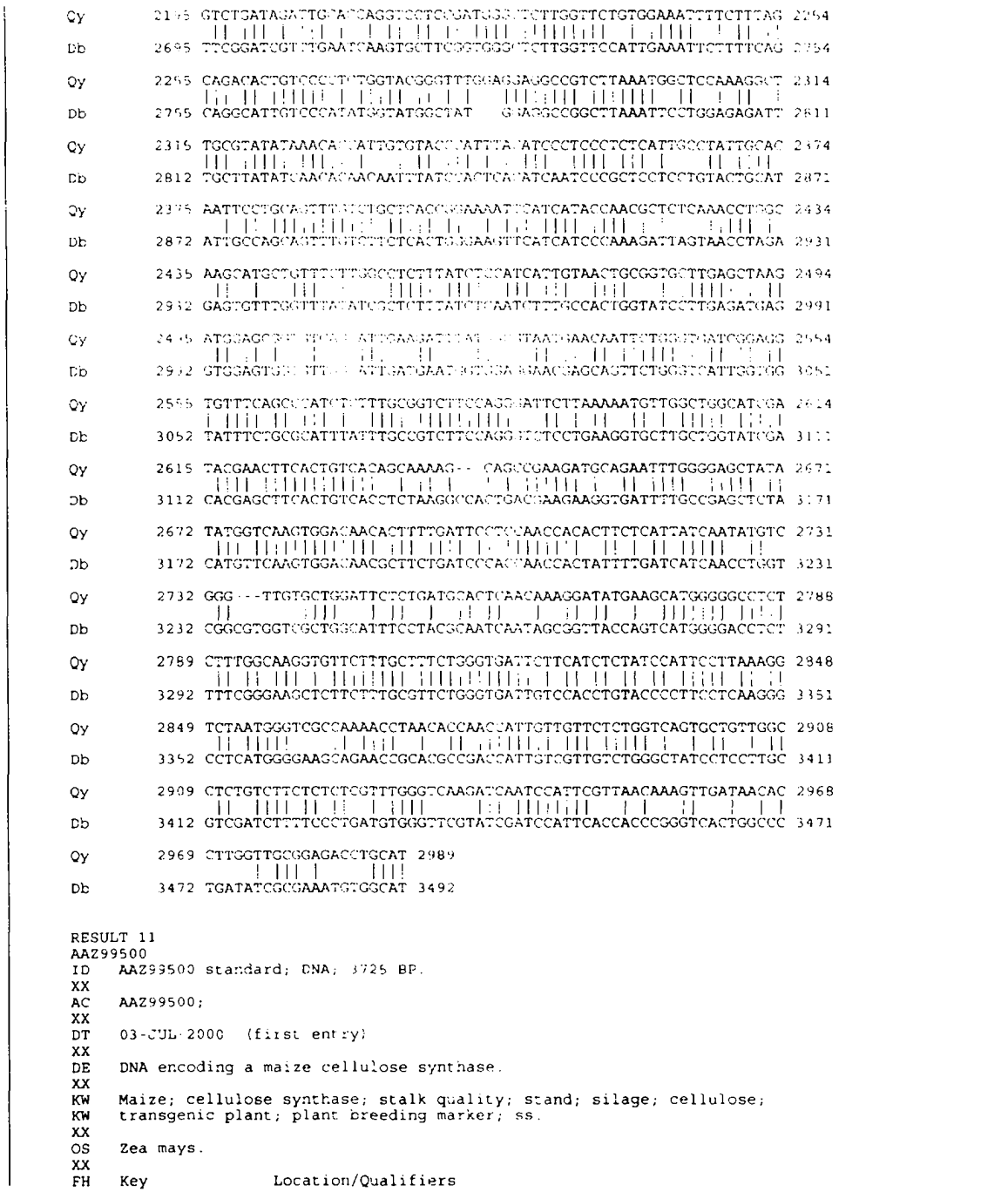


RESULT 10
AAZ99533
ID AAZ99533 standard; DNA; 3704 BP.
XX
AC AAZ99533;
XX
DT 03-JUL-2000 (first entry)
XX
DE DNA encoding a maize cellulose synthase.
XX
KW Maize; cellulose synthase; stalk quality; stand; silage; cellulose;
KW transgenic plant; plant breeding; marker; ss.
XX
OS Zea mays.
XX
FH Key Location/Qualifiers
FT CDS 272..3496
FT /*tag= a
FT /product= "cellulose synthase"
FT /transl_except= (pos:1544..1546,aa:Xaa)
FT /note= "no termination codon given, Xaa is an
FT unspecified amino acid"
XX
PN WO200009706-A2.
XX
PD 24-FEB-2000.
XX
PF 16-AUG-1999; 99WO-US18760.
XX
PR 17-AUG-1998; 98US-0096822.
XX
PA (PION-) PIONEER HI-BRED INT INC.
XX
PI Dhugga KS, Helentjaris TG, Bowen BA, Wang X;
XX
DR WPI; 2000-224343/19.
DR P-PSDB; AAY84121.
XX
PT New genes which encode maize cellulose synthase polypeptides in plants
PT useful for modulating the expression of cellulose synthase in plants
PT and to produce transgenic plants expressing the novel protein -
XX
PS Claim 1; Page 191-196; 119pp; English.
XX
CC The present sequence encodes a maize cellulose synthase polypeptide.
CC The cellulose synthase can be used for the improvement of stalk quality
CC for improved stand or silage. It also provides an increased concentration
CC of cellulose in the pericarp, hardening the kernel and improving its
CC handling ability. The sequences are used to produce transgenic plants
CC and seeds expressing the cellulose synthase. The polynucleotide is
CC used for modulating, preferably increasing, the level of the synthase
CC in a plant cell. The plants are preferably monocots. The polynucleotide
CC is also used as a probe or primer in the detection quantitation or
CC isolation of gene transcripts. The probes are useful in detecting
CC deficiencies in the level of mRNA in screenings for desired transgenic
CC plant, for detecting mutations in the gene, for monitoring upregulation
CC of expression or changes in enzyme activity in screening assays of
CC compounds, for detection of any number of allelic variants of the gene,
CC or for use as molecular markers in plant breeding programs. The
CC isolated nucleic acids of the present invention can also be used for
CC recombinant expression of their encoded polypeptides or for use as
CC immunogens in the preparation and/or screening of antibodies. The
CC proteins can be employed in assays for enzyme agonists or antagonists
CC of enzyme function or for use of immunogens or antigens to obtain
CC antibodies specifically immunoreactive with a protein.
XX
SQ Sequence 3704 BP; 909 A; 838 C; 962 G; 994 T; 1 other;

Query Match 35.2%; Score 1137.8; DB 21; Length 3704;
Best Local Similarity 67.8%; Pred. No. 3.2e-291;
Matches 1682; Conservative 0; Mismatches 784; Indels 15; Gaps 6;







```
RESULT 11
AAZ99500
ID   AAZ99500 standard; DNA; 3725 BP.
XX
AC   AAZ99500;
XX
DT   03-JUL-2000  (first entry)
XX
DE   DNA encoding a maize cellulose synthase.
XX
KW   Maize; cellulose synthase; stalk quality; stand; silage; cellulose;
KW   transgenic plant; plant breeding marker; ss.
XX
OS   Zea mays.
XX
FH   Key             Location/Qualifiers
```

```
FT   CDS             179..3398
FT                   /*tag= a
FT                   /product= "cellulose synthase"
FT                   /note= "no termination codon given"
XX
PN   WO200009706-A2.
XX
PD   24-FEB-2000.
XX
PF   16-AUG-1999;   99WO-US18760.
XX
PR   17-AUG-1998;   98US-0096822.
XX
PA   (PION ) PIONEER HI-BRED INT INC.
XX
PI   Dhugga KS, Helentjaris TG, Bowen BA, Wang X;
XX
DR   WPI; 2000-224343/19.
DR   P-PSDB; AAY84110.
XX
PT   New genes which encode maize cellulose synthase polypeptides in plants
PT   useful for modulating the expression of cellulose synthase in plants
PT   and to produce transgenic plants expressing the novel protein -
XX
PS   Claim 1; Page 105-110; 119pp; English.
XX
CC   The present sequence encodes a maize cellulose synthase polypeptide.
CC   The cellulose synthase can be used for the improvement of stalk quality
CC   for improved stand or silage. It also provides an increased concentration
CC   of cellulose in the pericarp, hardening the kernel and improving its
CC   handling ability. The sequences are used to produce transgenic plants
CC   and seeds expressing the cellulose synthase. The polynucleotide is
CC   used for modulating, preferably increasing, the level of the synthase
CC   in a plant cell. The plants are preferably monocots. The polynucleotide
CC   is also used as a probe or primer in the detection quantitation or
CC   isolation of gene transcripts. The probes are useful in detecting
CC   deficiencies in the level of mRNA in screenings for desired transgenic
CC   plant, for detecting mutations in the gene, for monitoring upregulation
CC   of expression or changes in enzyme activity in screening assays of
CC   compounds, for detection of any number of allelic variants of the gene,
CC   or for use as molecular markers in plant breeding programs. The
CC   isolated nucleic acids of the present invention can also be used for
CC   recombinant expression of their encoded polypeptides or for use as
CC   immunogens in the preparation and/or screening of antibodies. The
CC   proteins can be employed in assays for enzyme agonists or antagonists
CC   of enzyme function or for use of immunogens or antigens to obtain
CC   antibodies specifically immunoreactive with a protein.
XX
SQ   Sequence 3725 BP; 927 A; 807 C; 958 G; 1041 T, 0 other;

  Query Match             35.2%;  Score 1137.6;  DB 21;  Length 3725;
  Best Local Similarity   69.0%;  Pred. No. 3.6e 291;
  Matches 1662;  Conservative    0;  Mismatches  729;  Indels   17;  Gaps    7;

Qy        553 TTTATCCAATTCCACGCAACAAGCTCACACCATACAGAGCAGTGATCATTATGCGACTGG 612
              ||   |||||| |    ||| ||||||  |  ||| |   ||| |||| | || ||
Db        951 TTGTGCCAATTTCCTCAAACCAGCTCAACCTTTACCGGATAGTAATCATTCTCCGTCTTA 1010

Qy        613 TCATTCTGGGCCTCTTCTTCCACTTCAGAATAACAAATCCTGTCGATAGTGCCTTTGGCC 672
              |||| ||| || |||||||||| |  | || |  |||| .|  .| ||| | |||
Db       1011 TCATCCTGTGCTTCTTCTTCCAATATCGTATCAGTCATCCAGTGCGTAATGCTTATGGAT 1070

Qy        673 TGTGGCTTACTTCTGTCATATGTGAGATCTGGTTTGCATTTTCTTGGGTGTTGGATCAGT 732
              |||||    |||| || ||||| |||||||||| || || | |  | |||||||
Db       1071 TGTGGCTAGTATCTGTTATCTGTGAGGTCTGGTTTGCCTTGTCCTGGCTTCTAGATCAGT 1130

Qy        733 TCCCCAAGTGGAATCCTGTCAATAGAGAAACGTATATCGAAAGGCTGTCGGCAAGGTATG 792
              |||| || ||| |||| ||| | | || ||| ||| .|||  |    |||||||
Db       1131 TCCCAAAATGGTATCCAATCAACCGTGAGACATATCTCGACAGGCTTGCATTGAGGTATG 1190

Qy        793 AAAGAGAGGGTGAGCCTTCTCAGCTTGCTGGTGTGGATTTTTTCGTGAGTACTGTTGATC 852
              | |||||||| ||||| || |||| ||    | ||| | || || |||| || || |
Db       1191 ATAGAGAGGGAGAGCCATCACAGCTGGCTCCCATTGATGTCTTTGTCAGTACAGTGGATC 1250

Qy        853 CGCTGAAGGAACCGCCATTGATCACTGCCAATACAGTCCTTTCCATCCTTGCTGTGGACT 912
              |  |||||||||| ||| ||||||| ||||| || ||  | ||||| |||||||||| |
Db       1251 CATTGAAGGAACCTCCACTGATCACAGCCAACACTGTTTTGTCCATTCTTGCTGTGGATT 1310

Qy        913 ATCCCGTCGATAAAGTCTCCTGCTACGTGTCTGATGATGGTGCAGCTATGCTTTCATTTG 972
              | || || || ||||| || ||||| || |||||||||||  |||||||||  | ||||
Db       1311 ACCCTGTTGACAAAGTGTCATGCTATGTTTCTGATGATGGCTCAGCTATGCTGACTTTTG 1370

Qy        973 AATCTCTTGTAGAAACAGCTGAGTTTGCAAGGAAGTGGGTTCCGTTCTGCAAAAAATTCT 1032
              | |||||    ||||| || || ||||| || |||||||||| || || || ||   |
Db       1371 AGTCTCTCTCTGAAACTGCCGAATTTGCTAGAAAGTGGGTTCCCTTTTGTAAGAAGCACA 1430

Qy       1033 CAATTGAACCAAGAGCACCGGAGTTTTACTTCTCACAGAAAATTGATTACTTGAAAGACA 1092
                |||||||||||||| || || |||||||  | || ||||| |||||| |||| ||||
Db       1431 ATATTGAACCAAGAGCTCCAGAATTTTACTTTGCTCAAAAAATAGATTACCTGAAGGACA 1490

Qy       1093 AGGTTCAACCTTCTTTCGTGAAAGAACGTAGAGCAATGAAAAGGGATTATGAAGAGTACA 1152
              | |||||||||| || || || ||| |  |||||||||| || || |||||||| | ||
Db       1491 AAATTCAACCTTCATTTGTTAAGGAAAGACGAGCAATGAAGAGAGAGTATGAAGAATTCA 1550

Qy       1153 AAGTCCGAGTTAATGCCCTGGTAGCAAAGGCTCAGAAAACACCTGAAGAAGGATGGACTA 1212
              || |  || | |||||||| || || || || ||||||  |||||||| || ||||| |
Db       1551 AAATAAGAATCAATGCCCTTGTTGCCAAAGCACAGAAAGTGCCTGAAGAGGGGTGGACCA 1610

Qy       1213 TGCAAGATGGAACACCTTGGCCTGGGAATAACACACGTGATCACCCTGGGCATGATTCAG 1272
              ||   |||||||  |||||||||||||||||| |  | || || ||||| ||||||||||
Db       1611 TGGCTGATGGAACTGCTTGGCCTGGGAATAACCCTAGGGACCATCCTGG-CATGATTCAG 1669

Qy       1273 GTCTTCCTTGGGAAATACTGGGAGCTCGTGACATTGAAGGAAATGAACTACCTCGTCTAG 1332
              || ||| | ||| |      | | ||| ||| ||||| ||| |||||||| |||| ||||| |
Db       1670 GTGTTCTTGGGGCACAGTGGTGGGCT--TGACACTGATGGAAATGAATTACCACGTCTTG 1727

Qy       1333 TATATGTCTCCAGGGAGAAGAGACCTGGCTACCAGCACCACAAAAAGGCTGGTGCAGAAA 1392
              | ||||||||  | || |||||||| ||||  ||||| ||||| ||||||||||||   |
Db       1728 TCTATGTCTCTCGTGAAAAGAGACCAGGCTTTCAGCATCACAAGAAGGCTGGTGCAATGA 1787

Qy       1393 ATGCTCTGGTGAGAGTGTCTGCAGTACTCACAAATGCTCCCTACATCCTCAATGTTGATT 1452
              |||| ||| |  | || ||||| || || ||||||| | ||||  | |||||||| ||||
Db       1788 ATGCACTGATTCGTGTATCTGCTGTGCTGACAAATGGTGCCTATCTTCTCAATGTGGATT 1847

Qy       1453 GTGATCACTATGTAAACAATAGCAAGGCTGTTCGAGAGGCAATGTGCATCCTGATGGACC 1512
              |||| || ||  | || |  |||| ||| || |||| |||||||| || ||||||| |
Db       1848 GTGACCATTACTTCAATAGCAGCAAAGCTCTTAGAGAAGCAATGTGCTTCATGATGGATC 1907

Qy       1513 CACAAGTAGGTCGAGATGTATGCTATGTGCAGTTCCCTCAGAGGTTTGATGGCATAGATA 1572
              ||    |||| |  |    || ||||| || || || || || |||||||||| ||
Db       1908 CAGCTCTAGGAAGGAAAACTTGTTATGTACAATTTCCACAAAGATTTGATGGCATTGACT 1967

Qy       1573 AGAGTGATCGCTACGCCAATCGTAACGTAGTTTTCTTTGATGTTAACATGAAAGGGTTGG 1632
              |  ||||| || || ||| | ||| |||| ||||||||| | ||||||||||  | |
Db       1968 TGCACGATCGATATGCTAATAGGAACATAGTCTTCTTTGATATCAACATGAAAGGTCTAG 2027

Qy       1633 ATGGCATTCAAGGACCAGTATACGTAGGAACTGGTTGTGTTTTCAACAGGCAAGCACTTT 1692
              |||||||||| || |||| || || |||||  || ||   ||||| ||||| ||  | |
Db       2028 ATGGCATTCAGGGTCCAGTCTATGTGGGAACAGGATGCTGTTTCAATAGGCAGGCTTTGT 2087

Qy       1693 ACGGCTACGGGCCTCCTTCTATGCCCAGCTTACGCAAGAGAAAGGATTCTTCATCCTGCT 1752
              | || || |  ||| ||  |  |  |  ||  | |  |||
Db       2088 ATGGATATGATCCTGTTTTGACTGAAGCTGATCTGGAACCTAACATTGTTGTTAAGAGCT 2147

Qy       1753 TCTCATGTTGCTGCCCCTCAAAGAAGAAGCCTGCTCAAGAT----CCAGCTGAGGTATAC 1808
              ||  || |  |       |||||| ||| |  |  |||       ||| |   |||
Db       2148 GCTGTGGTAGAAGGAAGAGAAAGAACAAGAGTTATATGGATAGTCAAAGCCGTATTATGA 2207

Qy       1809 AGAGATGCAAAAAGAGAGGATCTCAATGCTGCCATATTTAATCTTACAGAGATTGATAAT 1868
              ||||| || ||       |  || ||     |  |  |  ||| |  | |
Db       2208 AGAGA-ACAGAATCTTCAGCTCCCATCTTTAACATGGAAGACATCGAGGAGGGTATTGAA 2266

Qy       1869 TATGACGAGCATGAAAGGTCAATGCTGATCTCCCAGTTGAGCTTTGAGAAAACTTTTGGC 1928
              |  ||| |||||||||| |||| || ||||||  ||  || ||||||   |||||
Db       2267 GGTTATGAGGATGAAAGGTCAGTGCTTATGTCCCAGAGGAAATTGGAGAAACGCTTTGGT 2326
```

```
y       1929 TTATCTTCTGTCTTCATTGAGTCTACACTAATGGAGAATGGAGGAGTACCCGAGTCTGCC 1988
                |||| |  |||||||||  || ||  | |||    | || ||  ||||    || .
b       2327 CAGTCTCCAATCTTCATTGCATCCACCTTTATGACTCAAGGTGGCATACCACCTTCAACA 2386

y       1989 AACTCACCACCATTCATCAAGGAAGCGATTCAAGTCATCGGCTGTGGCTATGAAGAGAAG 2048
                ||| || |  |  |  | |||||||| || || || ||| ||||||| || || || ||
b       2387 AACCCAGCTTCTCTACTGAAGGAAGCTATCCATGTTATCAGCTGTGGGTACGAGGACAAA 2446

y       2049 ACTGAATGGGGAAAACAGATTGGTTGGATATATGGGTCAGTCACTGAGGATATCTTAAGT 2108
                |||||||||||||||| ||||||| ||||| ||||| ||||| || ||||||||  | | |
b       2447 ACTGAATGGGGAAAAGAGATTGGCTGGATCTATGGTTCAGTTACAGAGGATATTCTGACT 2506

y       2109 GGCTTCAAGATGCACTGCCGAGGATGGAGATCAATTTACTGCATGCCCGTAAGGCCTGCA 2168
                || || || |||||     |||| |||  |||||| ||||||||||    | | |||
b       2507 GGGTTTAAAATGCATGCAAGAGGCTGGCAATCAATCTACTGCATGCCACCACGACCTTGT 2566

y       2169 TTCAAAGGATCTGCACCCATCAACCTGTCTGATAGATTGCACCAGGTCCTCCGATGGGCT 2228
                ||||| || |||||||| ||||| || |||||| |   |  | |||||  ||||| ||||||
b       2567 TTCAAGGGTTCTGCACCAATCAATCTTTCTGATCGTCTTAATCAGGTGCTCCGTTGGGCT 2626

y       2229 CTTGGTTCTGTGGAAATTTTCTTTAGCAGACACTGTCCCCTCTGGTACGGGTTTGGAGGA 2288
                ||||| || |||||||||  |  |||||||||| |||||  | |||||  || |
b       2627 CTTGGGTCAGTGGAAATTCTGCTTAGCAGACATTGTCCTATATGGTATGGCT---ACAAT 2683

y       2289 GGCCGTCTTAAATGGCTCCAAAGGCTTGCGTATATAAACACCATTGTGTACCCATTTACA 2348
                || ||  | ||      |  | |||| || || || ||||||||| || ||| ; |||
b       2684 GGGCGATTGAAGCTTTTGGAGAGGCTGGCTTACATTAACACCATTGTTTATCCAATCACA 2743

y       2349 TCCCTCCCTCTCATTGCCTATTGCACAATTCCTGCAGTTTGTCTGCTCACCGGAAAATTC 2408
                ||  | || || || ||||||||      |||||||  | |||||  || |||   |||||
b       2744 TCTGTTCCGCTTATCGCCTATTGTGTGCTTCCTGCTATCTGTCTTCTTACCAATAAATTT 2803

y       2409 ATCATACCAACGCTCTCAAACCTGGCAAGCATGCTGTTTCTTGGCCTCTTTATCTCCATC 2468
                ||||| ||    | |    ||   ||  | ||| | ||  ||   || |||  ||||||
b       2804 ATCATTCCTGAGATTAGTAATTATGCTGGAATGTTCTTCATTCTTCTTTTTGCCTCCATT 2863

y       2469 ATTGTAACTGCGGTGCTTGAGCTAAGATGGAGCGGTGTCAGCATTGAAGATTTATGGCGT 2528
                 | | |||||    |  | |||||  |||||||| |||||  ||||||||||||  ||| |
b       2864 TTCGCAACTGGTATATTGGAGCTCAGATGGAGTGGTGTTGGCATTGAAGATTGGTGGAGA 2923

y       2529 AATGAACAATTCTGGGTGATCGGAGGTGTTTCAGCCCATCTCTTTGCGGTCTTCCAGGGA 2588
                .|||| || || ||||| || || ||    || ||||||||||| ||||| ||||||||
b       2924 AATGAGCAGTTTTGGGTTATTGGTGGCACCTCTGCCCATCTCTTCGCGGTGTTCCAGGGT 2983

y       2589 TTCTTAAAAATGTTGGCTGGCATCGATACGAACTTCACTGTCACAGCAAAAGCA---GCC 2645
                |  | ||| |||||||||| || ||||| ||||||| || ||  |||| |||   |
b       2984 CTGCTGAAAGTGTTGGCTGGGATTGATACCAACTTCACAGTTACCTCAAAGGCATCTGAT 3043

y       2646 GAAGATGCAGAATTTGGGGAGCTATATATGGTCAAGTGGACAACACTTTTGATTCCTCCA 2705
                || ||||  || ||||  ||||||||| || |||||||||| |   |  | || ||||
b       3044 GAGGATGGCGACTTTGCTGAGCTATATGTGTTCAAGTGGACCAGTTTGCTCATCCCTCCG 3103

y       2706 ACCACACTTCTCATTATCAATATGTCGGG---TTGTGCTGGATTCTCTGATGCACTCAAC 2762
                |||||  ||||  | || ||  ||   ||       || ||| | ||  ||||  | |||
b       3104 ACCACTGTTCTTGTCATTAACCTGGTCGGAATGGTGGCAGGAATTTCGTATGCCATTAAC 3163

y       2763 AAAGGATATGAAGCATGGGGGCCTCTCTTTGGCAAGGTGTTCTTTGCTTTCTGGGTGATT 2822
                |  || ||  || | ||||| || |||||||| ||| |||||||  |  ||||||||||
b       3164 AGCGGCTACCAATCCTGGGGTCCGCTCTTTGGAAAGCTGTTCTTCTCGATCTGGGTGATC 3223

y       2823 CTTCATCTCTATCCATTCCTTAAAGGTCTAATGGGTCGCCAAAACCTAACACCAACCATT 2882
                || |||||||| || ||||| || ||||| |||||  | || ||||  || ||||| ||
b       3224 CTCCATCTCTACCCCTTCCTCAAGGGTCTCATGGGCAGGCAGAACCGCACGCCAACAATC 3283

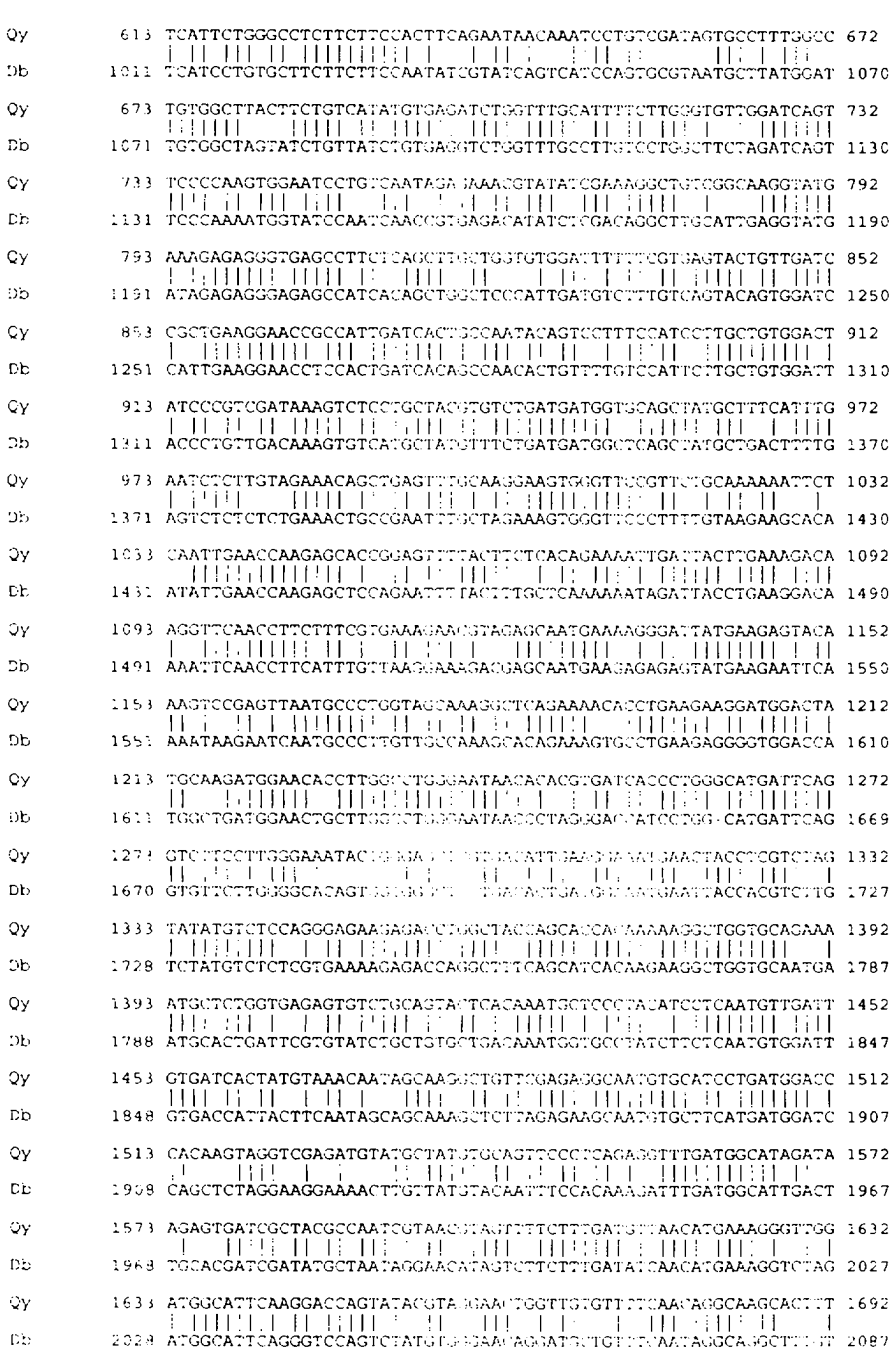
y       2883 GTTGTTCTCTGGTCAGTGCTGTTGGCCTCTGTCTTCTCTCTCGTTTGGGTCAAGATCAAT 2942
                ||  |  | |||||  | ||  | || ||| |||||||  |  | |||||  ||||||  ||
b       3284 GTCATCGTTTGGTCCATCCTCCTTGCGTCTATCTTCTCCTTGCTGTGGGTGAAGATCGAT 3343

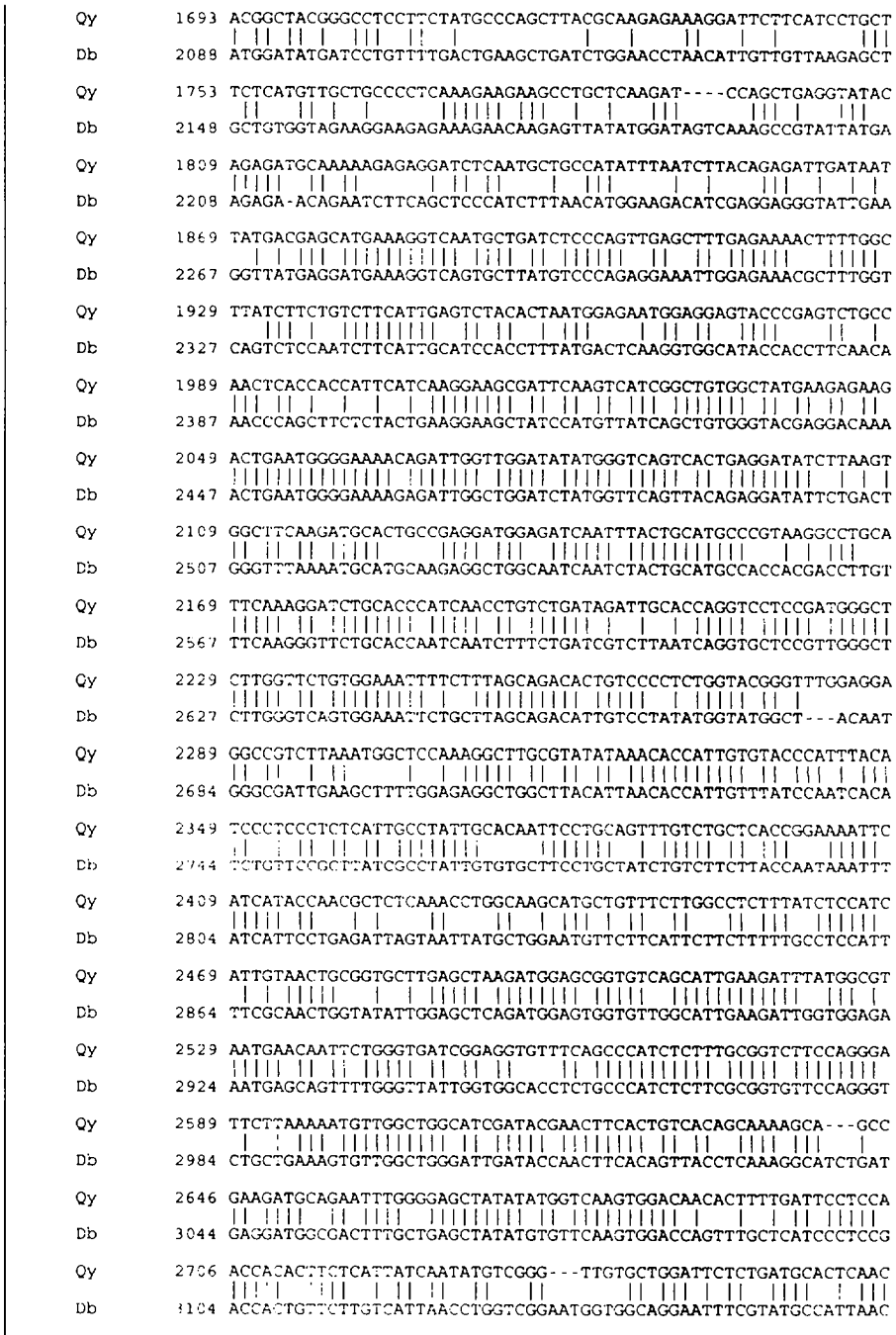
y       2943 CCATTCGT 2950
                || ||| |
b       3344 CCTTTCAT 3351
```

```
RESULT 12
AAZ99506
ID   AAZ99506 standard; DNA; 3725 BP.
XX
AC   AAZ99506;
XX
DT   03-JUL-2000  (first entry)
XX
DE   DNA encoding a maize cellulose synthase.
XX
KW   Maize; cellulose synthase; stalk quality; stand; silage; cellulose;
KW   transgenic plant; plant breeding marker; cs.
XX
OS   Zea mays.
XX
FH   Key             Location/Qualifiers
FT   CDS             179..3376
FT                   /*tag= a
FT                   /product= "cellulose synthase"
FT                   /note= "no termination codon given"
XX
PN   WO200039706-A2.
XX
PD   24-FEB-2000.
XX
PF   16-AUG-1999;   99WO-US018768.
XX
PR   17-AUG-1998;   98US-0096422.
XX
PA   (PION-) PIONEER HI BRED INT INC.
XX
PI   Dhugga KS, Helentjaris TG, Bowen BA, Wang X;
XX
DR   WPI; 2000-224343/19.
DR   P-PSDB; AAY84112.
XX
PT   New genes which encode maize cellulose synthase polypeptides in plants
PT   useful for modulating the expression of cellulose synthase in plants
PT   and to produce transgenic plants expressing the novel protein.
XX
PS   Claim 1; Page 121-126; 119pp; English
XX
CC   The present sequence encodes a maize cellulose synthase polypeptide.
CC   The cellulose synthase can be used for the improvement of stalk quality
CC   for improved stand or silage. It also provides an increased concentration
CC   of cellulose in the pericarp, hardening the kernel and improving its
CC   handling ability. The sequences are used to produce transgenic plants
CC   and seeds expressing the cellulose synthase. The polynucleotide is
CC   used for modulating, preferably increasing, the level of the synthase
CC   in a plant cell. The plants are preferably monocots. The polynucleotide
CC   is also used as a probe or primer in the detection quantitation or
CC   isolation of gene transcripts. The probes are useful in detecting
CC   deficiencies in the level of mRNA in screenings for desired transgenic
CC   plant, for detecting mutations in the gene, for monitoring upregulation
CC   of expression or changes in enzyme activity in screening assays of
CC   compounds, for detection of any number of allelic variants of the gene,
CC   or for use as molecular markers in plant breeding programs. The
CC   isolated nucleic acids of the present invention can also be used for
CC   recombinant expression of their encoded polypeptides or for use as
CC   immunogens in the preparation and/or screening of antibodies. The
CC   proteins can be employed in assays for enzyme agonists or antagonists
CC   of enzyme function or for use of immunogens or antigens to obtain
CC   antibodies specifically immunoreactive with a protein.
XX
SQ   Sequence 3725 BP; 927 A; 807 C; 950 G; 1041 T; 0 other;

  Query Match           35.2%;  Score 1137.6;  DB 21;  Length 3725;
  Best Local Similarity   69.0%;  Pred. No. 3.6e-291;
  Matches 1662;  Conservative    0;  Mismatches  729;  Indels   17;  Gaps    7;

Qy        553 TTTATCCAATTCCACGCAACAAGCTCACACCATACAGAGCAGTGATCATTATGCGACTGG 612
                ||   |||||| |    ||| ||||.|  |  ||| |   ||| |||||| | || ||
Db        951 TTGTGCCAATTTCCTCAAACCAGCTCAACCTTTACCGGATAGTAATCATTCTCCGTCTTA 1010
```





```
      3164 AGCGGCTACCAATCCTGGGGTCCGCTCTTTGGAAAGCTGTTCTTCTCGATCTGGGTGATC 3223

      2823 CTTCATCTCTATCCATTCCTTAAAGGTCTAATGGGTCGCCAAAACCTAACACCAACCATT 2882

      3224 CTCCATCTCTACCCCTTCCTCAAGGGTCTCATGGGCAGGCAGAACCGCACGCCAACAATC 3283

      2883 GTTGTTCTCTGGTCAGTGCTGTTGGCCTCTGTCTTCTCTCTCGTTTGGGTCAAGATCAAT 2942

      3284 GTCATCGTTTGGTCCATCCTCCTTGCGTCTATCTTCTCCTTGCTGTGGGTGAAGATCGAT 3343

      2943 CCATTCGT 2950

      3344 CCTTTCAT 3351


SULT 13
AZ99521
)  AAZ99521 standard; DNA; 3725 BP.
(
:  AAZ99521;
(
-  03-JUL-2000  (first entry)
(
:  DNA encoding a maize cellulose synthase.
(
W  Maize; cellulose synthase; stalk quality; stand; silage; cellulose;
W  transgenic plant; plant breeding marker; ss.
(
3  Zea mays.
(
F  Key             Location/Qualifiers
F  CDS             179..3397
F                  /*tag= a
F                  /product= "cellulose synthase"
F                  /note= "no termination codon given"
(
W  WO200009706-A2.
K
D  24-FEB-2000.
K
F  16-AUG-1999;   99WO-US18760.
K
R  17-AUG-1998;   98US-0096822.
K
A  (PION-) PIONEER HI-BRED INT INC.
X
I  Dhugga KS,  Helentjaris TG,  Bowen BA,  Wang X;
X
R  WPI; 2000-224343/19.
R  P-PSDB; AAY84117.
X
T  New genes which encode maize cellulose synthase polypeptides in plants
T  useful for modulating the expression of cellulose synthase in plants
T  and to produce transgenic plants expressing the novel protein -
X
S  Claim 1; Page 160-165; 119pp; English.
X
C  The present sequence encodes a maize cellulose synthase polypeptide.
C  The cellulose synthase can be used for the improvement of stalk quality
C  for improved stand or silage. It also provides an increased concentration
C  of cellulose in the pericarp, hardening the kernel and improving its
C  handling ability. The sequences are used to produce transgenic plants
C  and seeds expressing the cellulose synthase. The polynucleotide is
C  used for modulating, preferably increasing, the level of the synthase
C  in a plant cell. The plants are preferably monocots. The polynucleotide
C  is also used as a probe or primer in the detection quantitation or
C  isolation of gene transcripts. The probes are useful in detecting
C  deficiencies in the level of mRNA in screenings for desired transgenic
C  plant, for detecting mutations in the gene, for monitoring upregulation
C  of expression or changes in enzyme activity in screening assays of
C  compounds, for detection of any number of allelic variants of the gene,
C  or for use as molecular markers in plant breeding programs. The
C  isolated nucleic acids of the present invention can also be used for
C  recombinant expression of their encoded polypeptides or for use as
C  immunogens in the preparation and/or screening of antibodies. The
C  proteins can be employed in assays for enzyme agonists or antagonists
C  of enzyme function or for use of immunogens or antigens to obtain
C  antibodies specifically immunoreactive with a protein.
XX
SQ  Sequence 3725 BP; 927 A; 807 C; 950 G; 1041 T; 0 other;

  Query Match             85.2%;  Score 1137.6;  DB 21;  Length 3725;
  Best Local Similarity   69.0%;  Pred. No. 3.6e-291;
  Matches 1662;  Conservative    0;  Mismatches  728;  Indels   17;  Gaps    7;

Qy        553 TTTATCCAATTCCAATGCAACAAGCTCACAGCATACAGAGCAGTGATCATTATGCGACTGG 612
Db        961 TTGTGCCAATTTC TTAAACCAGCTCAACTTTACGGGATAGTAATCATTCTCCGTCTTA 1019

Qy        613 TCATTCTGGGCCTCTTCTTCCACTTCAGAATAACAAATCCTGTCGATAGTGCCTTTGGCC 672
Db       1020 TCATCCTCTCTTTC TTTAATA TCATTAGTCATCCAGTGCGTAATGCTTATGGAT 1078

Qy        673 TGTAGCTTA TC  TATACTTA CC  TTTTC ATTTTCTTGG TGTTGGATCAGT 732
Db       1079 TGTGGCTAATAT TCTATCTTCTA  TTGGTCCTGCTTCTAGATCAGT 1130

Qy        733 TCCCCAAGTGGAATCCTGTCAATAGAGAAACGTATATCGAAAGGCTGTCGGCAAGGTATG 792
Db       1131 TCCCAAAATGGTATCCAATCAACCGTGAGACATATCTCGACAGGCTTGCATTGAGGTATG 1190

Qy        793 AAAGAGAGGGTGAGTCTTCTCAGCTTGCTGGGTGGATTTTTTCGTGAGTACTGTTGATC 852
Db       1191 ATAGAGAGGGAGAGTCATCACAGCTGGTTCCATTGATGTCTTTGTCAGTACAGTGGATC 1250

Qy        853 CGCTGAAGGAACCGTCATTGATCACTGCCAATACAGTTCTTTCCATCCTTGCTGTGGACT 912
Db       1251 CATTGAAGGAACCTTCACTGATCACACGTAACACTGTTTTGTCCATTCTTGCTGTGGACT 1310

Qy        913 ATCCCGTCGATAAAGTCTCCTGCTACGTGTCTGATGATGGTGCAGCTATGCTTTCATTTG 972
Db       1311 ACCCTGTTGACAAAGTGTCATGCTATGTTTCTGATGATGGCTCAGCTATGCTGACTTTTG 1370

Qy        973 AATCTCTTGTAGAAACAGCTGAGTTTGCAAGGAAGTGGGTTCCGTTCTGCAAAAAATTCT 1032
Db       1371 AGTCTCTCCCTGAAACTCCGAATTTGCTAGAAAGTGGGTTCCCTTTTGTAAGAAGCACA 1430

Qy       1033 CAATTGAACCAAGAGCACCGGAGTTTTACTTTCTCACAGAAAATTGATTACTTGAAAGACA 1092
Db       1431 ATATTGAACCAAGAGCTCCAGAATTTTACTTTGCTCAAAAAATAGATTACCTGAAGGACA 1490

Qy       1093 AGGTTCAACCTTCTTTCGTGAAAGAACGTAGAGCAATGAAAAGGGATTATGAAGAGTACA 1152
Db       1491 AAATTCAACCTTCATTTGTTAAGGAAAGAAGAGCAATGAAGAGAGAGTATGAAGAATTCA 1550

Qy       1153 AAGTCCGAGTTAATGCCCTGGTAGCAAAGGCTCAGAAAACACCTGAAGAAGGATGGACTA 1212
Db       1551 AAATAAGAATCAATGCCCTTGTTGCCAAAGCACAGAAAGTGCCTGAAGAGGGGTGGACCA 1610

Qy       1213 TGCAAGATGGAACACTTGGCCTGGGAATAACACACGTGATCACCCTGGGCATGATTCAG 1272
Db       1611 TGGCTGATGGAACTGCTTGGCCTGGGAATAACCCTAGGGACCATCCTGG-CATGATTCAG 1669

Qy       1273 GTCTTCCTTGGGAAATACTGGGAGCTGGTGACATTGAAGGAAATGAACTACCTCGTCTAG 1332
Db       1670 GTGTTCTTGGGGCAAAGTGGTGGGCT--TGACACTGATGGAAATGAATTACCACGTCTTG 1727

Qy       1333 TATATGTCTTCAGGGAGAAGAGACCTGGCTACCAGCACCACAAAAAGGCTGGTGCAGAAA 1392
Db       1728 TCTATGTCTCTCGTGAAAAGAGACCAGGCTTTCAGCATCACAAGAAGGCTGGTGCAATGA 1787

Qy       1393 ATGCTCTGGTGAGAGTGTCTGCAGTACTCACAAATGCTCCCTACATCCTCAATGTTGATT 1452
Db       1788 ATGCACTGATTCGTGTATCTGCTGTGCTGACAAATGGTGCCTATCTTCTCAATGTGGATT 1847
```

```
Qy       1453 GTGATCACTATGTAAACAATAGCAAGGCTGTTCGAGAGGCAATGTGCATCCTGATGGACC 1512
Db       1848 GTGACCATTACTTCAATAGCAGCAAAGCTCTTAGAGAAGCAATGTGCTTCATGATGGATC 1907

Qy       1513 CACAAGTAGGTCGAGATGTATGCTATGTGCAGTTCCCTCAGAGGTTTGATGGCATAGATA 1572
Db       1908 CAGCTCTAGGAAGGAAAACTTGTTATGTACAATTTCCACAAAGATTTGATGGCATTGACT 1967

Qy       1573 AGAGTGATCGCTACGCCAATCGTAACGTAGTTTTCTTTGATGTTAACATGAAAGGGTTGG 1632
Db       1968 TGCACGATCGATATGCTAATAGGAACATAGTCTTCTTTGATATCAACATGAAAGGTCTAG 2027

Qy       1633 ATGGCATTCAAGGACCAGTATACGTAGGAACTGGTTGTGTTTTCAACAGGCAAGCACTTT 1692
Db       2028 ATGGCATTCAGGGTCCAGTCTATGTGGGAACAGGATGCTGTTTCAATAGGCAGGCTTTGT 2087

Qy       1693 ACGGCTACGGGGCCTCCTTCTATGCCCAGCTTACGCAAGAGAAAGGATTCTTCATCCTGCT 1752
Db       2088 ATGGATATGATCCTGTTTTGACTGAAGCTGATCTGGAACCTAACATTGTTGTTAAGAGCT 2147

Qy       1753 TCTCATGTTGCTGCCCCTCAAAGAAGAAGCCTGCTCAAGAT---CCAGCTGAGGTATAC 1808
Db       2148 GCTGTGGTAGAAGGAAGAGAAAGAACAAGAGTTATATGGATAGTCAAAGCCGTATTATGA 2207

Qy       1809 AGAGATGCAAAAAGAGAGGATCTCAATGCTGCCATATTTAATCTTACAGAGATTGATAAT 1868
Db       2208 AGAGA-ACAGAATCTTCAGCTCCCATCTTTAACATGGAAGACATCGAGGAGGGTATTGAA 2266

Qy       1869 TATGACGAGCATGAAAGGTCAATGCTGATCTCCCAGTTGAGCTTTGAGAAAACTTTTGGC 1928
Db       2267 GGTTATGAGGATGAAAGGTCAGTGCTTATGTCCCAGAGGAAATTGGAGAAACGCTTTGGT 2326

Qy       1929 TTATCTTCTGTCTTCATTGAGTCTACACTAATGGAGAATGGAGGAGTACCCGAGTCTGCC 1988
Db       2327 CAGTCTCCAATCTTCATTGCATCCACCTTTATGACTCAAGGTGGCATACCACCTTCAACA 2386

Qy       1989 AACTCACCACCATTCATCAAGGAAGCGATTCAAGTCATCGGCTGTGGCTATGAAGAGAAG 2048
Db       2387 AACCCAGCTTCTCTACTGAAGGAAGCTATCCATGTTATCAGCTGTGGGTACGAGGACAAA 2446

Qy       2049 ACTGAATGGGGAAAACAGATTGGTTGGATATATGGGTCAGTCACTGAGGATATCTTAAGT 2108
Db       2447 ACTGAATGGGGAAAAGAGATTGGCTGGATCTATGGTTCAGTTACAGAGGATATTCTGACT 2506

Qy       2109 GGCTTCAAGATGCACTGCCGAGGATG[illegible]CTTTACTGCATGCCCGTAAGGCCTGCA 2168
Db       2507 GGGTTTAAAATGCATGCAAGAG[illegible]CTACTGCATGCTACGACGTTTGT 2566

Qy       2169 TTCAAAGGATCTGCACCCATCAA[illegible]TCCTCCGATGGGCT 2228
Db       2567 TTCAAGGGTTCTGCACCAATCAATCT[illegible]TTAATCAGGTGCTCCGTTGGGCT 2626

Qy       2229 CTTGGTTCTGTGGAAATTTTCTTTAGCAGACACTGTCCCCTCTGGTACGGGTTTGGAGGA 2288
Db       2627 CTTGGGTCAGTGGAAATTCTGCTTAGCAGACATTGTCCTACATGGTATGGCT---ACAAT 2683

Qy       2289 GGCCGTCTTAAATGGCTCCAAAGGCTTGCGTATATAAACA[illegible]CATTGTGTACCCATTTACA 2348
Db       2684 GGGCGATTGAAGCTTTTGGAGAGGCTGGCTTACATTAACACCATTGTTTATCCAATCACA 2743

Qy       2349 TCCCTCCCTCTCATTGCCTATTGCACAATTGCTGCAGTTTGTCTGCTCACCGGAAAATTC 2408
Db       2744 TCTGTTCCGCTTATCGCCTATTGTGTGCTTCCTGCTATCTGTCTTCTTACCAATAAATTT 2803

Qy       2409 ATCATACCAACGCTCTCAAACCTGGCAAGCATGCTGTTTCTTGGCCTCTTTATCTCCATC 2468
Db       2804 ATCATTCCTGAGATTAGTAATTATGCTGGAATGTTCTTCATTGTTCTTTTTGCCTCCATT 2863

Qy       2469 ATTGTAACTGCGGTGCTTGAGCTAAGATGGAGCGGTGTCAG[illegible]ATTGAAGATTTATGGCGT 2528
Db       2864 TTCGCAACTGGTATATTGGAGCTCAGATGGAGTGGTGTTGGCATTG[illegible]AGATTGGTGGAGA 2923

Qy       2529 AATGAACAATTCTGGGTGATCGGA[illegible]TTCAGCCCATCT[illegible]GGTCTTCCAG[illegible]A 2588

                 ||||| || || ||||| || || ||    || |||||||||| ||||| ||||||||
Db       2924 AATGAGCAGTTTTGGGTTATTGGTGGCACCTCTGCCCATCTCTTCGCGGTGTTCCAGGGT 2983

Qy       2589 TTCTTAAAAAATGTTGGCTGGCATCGATACGAACTTCACTGTCACAGCAAAAGCA---GCC 2645
Db       2984 CTGCTGAAAGTGTTGGCTGGGATTGATACCAACTTCACAGTTACCTCAAAGGCATCTGAT 3043

Qy       2646 GAAGATGCAGAATTTGGGGAGCTATATATGGTCAAGTGGACAACACTTTTGATTCCTCCA 2705
Db       3044 GAGGATGGCGACTTTGCTGAGCTATATGTGTTCAAGTGGACCAGTTTGCTCATCCCTCCG 3103

Qy       2706 ACCACACTTCTCATTATCAATATGTCGGG---TTGTGCTGGATTCTCTGATGCACTCAAC 2762
Db       3104 ACCACTGTTCTTGTCATTAACCTGGTCGGAATGGTGGCAGGAATTTCGTATGCCATTAAC 3163

Qy       2763 AAAGGATATGAAGCATGGGGGCCTCTCTTTGGCAAGGTGTTCTTTGCTTTCTGGGTGATT 2822
Db       3164 AGCGGCTACCAATCCTGGGGTCCGCTCTTTGGAAAGCTGTTCTTCTCGATCTGGGTGATC 3223

Qy       2823 CTTCATCTCTATCCATTCCTTAAAGGTCTAATGGGTCGCCAAAACCTAACACCAACCATT 2882
Db       3224 CTCCATCTCTACCCCTTCCTCAAGGGTCTCATGGGCAGGCAGAACCGCACGCCAACAATC 3283

Qy       2883 GTTGTTCTCTGGTCAGTGCTGTTGGCCTCTGTCTTCTCTCTCGTTTGGGTCAAGATCAAT 2942
Db       3284 GTCATCGTTTGGTCCATCCTCCTTGCGTCTATCTTCTCCTTGCTGTGGGTGAAGATCGAT 3343

Qy       2943 CCATTCGT 2950
Db       3344 CCTTTCAT 3351
```

RESULT 14
AAC49550
ID AAC49550 standard; DNA; 3198 BP.
XX
AC AAC49550;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 61572.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB 2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.

06-MAY-1999; 99US-0132486.
06-MAY-1999; 99US-0132487.
07-MAY-1999; 99US-0132863.
11-MAY-1999; 99US-0134256.
14-MAY-1999; 99US-0134218.
14-MAY-1999; 99US-0134219.
14-MAY-1999; 99US-0134221.
14-MAY-1999; 99US-0134370.
18-MAY-1999; 99US-0134768.
19-MAY-1999; 99US-0134941.
20-MAY-1999; 99US-0135124.
21-MAY-1999; 99US-0135353.
24-MAY-1999; 99US-0135629.
25-MAY-1999; 99US-0136021.
27-MAY-1999; 99US-0136392.
28-MAY-1999; 99US-0136782.
01-JUN-1999; 99US-0137222.
03-JUN-1999; 99US-0137528.
04-JUN-1999; 99US-0137502.
07-JUN-1999; 99US-0137724.
08-JUN-1999; 99US-0138094.
10-JUN-1999; 99US-0138540.
10-JUN-1999; 99US-0138847.
14-JUN-1999; 99US-0139119.
16-JUN-1999; 99US-0139452.
16-JUN-1999; 99US-0139453.
17-JUN-1999; 99US-0139492.
18-JUN-1999; 99US-0139454.
18-JUN-1999; 99US-0139455.
18-JUN-1999; 99US-0139456.
18-JUN-1999; 99US-0139457.
18-JUN-1999; 99US-0139458.
18-JUN-1999; 99US-0139459.
18-JUN-1999; 99US-0139460.
18-JUN-1999; 99US-0139461.
18-JUN-1999; 99US-0139462.
18-JUN-1999; 99US-0139463.
18-JUN-1999; 99US-0139750.
18-JUN-1999; 99US-0139763.
21-JUN-1999; 99US-0139817.
22-JUN-1999; 99US-0139899.
23-JUN-1999; 99US-0140353.
23-JUN-1999; 99US-0140354.
24-JUN-1999; 99US-0140695.
28-JUN-1999; 99US-0140823.
29-JUN-1999; 99US-0140991.
30-JUN-1999; 99US-0141287.
01-JUL-1999; 99US-0141842.
01-JUL-1999; 99US-0142154.
02-JUL-1999; 99US-0142055.
06-JUL-1999; 99US-0142390.
08-JUL-1999; 99US-0142803.
09-JUL-1999; 99US-0142920.
12-JUL-1999; 99US-0142977.
13-JUL-1999; 99US-0143542.
14-JUL-1999; 99US-0143624.
15-JUL-1999; 99US-0144005.
16-JUL-1999; 99US-0144085.
16-JUL-1999; 99US-0144086.
19-JUL-1999; 99US-0144325.
19-JUL-1999; 99US-0144331.
19-JUL-1999; 99US-0144332.
19-JUL-1999; 99US-0144333.
19-JUL-1999; 99US-0144334.
19-JUL-1999; 99US-0144335.
20-JUL-1999; 99US-0144352.
20-JUL-1999; 99US-0144632.
20-JUL-1999; 99US-0144884.
21-JUL-1999; 99US-0144814.
21-JUL-1999; 99US-0145086.
21-JUL-1999; 99US-0145088.
22-JUL-1999; 99US-0145085.
22-JUL-1999; 99US-0145087.

PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192
PR 23-JUL-1999; 99US-0145145
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145916.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147483.
PR 09-AUG-1999; 99US-0147487.
PR 10-AUG-1999; 99US-0147571.
PR 11-AUG-1999; 99US-0148519.
PR 12-AUG-1999; 99US-0148541.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP 1999; 99US-0156458.
PR 29-SEP 1999; 99US-0156596.
PR 04-OCT 1999; 99US-0157117.
PR 05-OCT 1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT 1999; 99US-0159294.
PR 13-OCT 1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.

PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 35.1%; Score 1135.4; DB 21; Length 3198;
Best Local Similarity 68.1%; Pred. No. 1.3e-290;
Matches 1692; Conservative 0; Mismatches 761; Indels 30; Gaps 7;

```
Qy       519 CCGTCTGCAGAGGCTTCGGAGCCGCTTCAATTGTTTATCCAATTCCACGCAACAAGCTC 578
Db       706 CTGAATGACGAAGCGAGGCAGCCTCTGTCAAGGAAAGTTTCAATTCCTTCATCACGGATC 765

Qy       579 ACACCATACAGAGCAGTGATCATTATGCGACTGGTCATTCTGGGCCTCTTCTTCCACTTC 638
Db       766 AATCCTTACAGAATGGTTATTATGCTGCGGCTTGTTATCCTTTGTCTCTTCTTGCATTAC 825

Qy       639 AGAATAACAAATCCTGTCGATAGTGCCTTTGGCCTGTGGCTTACTTCTGTCATATGTGAG 698
Db       826 CGTATAACAAACCCAGTGCCAAATGCCTTTGCTCTATGGCTGGTCTCTGTGATATGTGAG 885

Qy       699 ATCTGGTTTGCATTTTCTTGGGTGTTGGATCAGTTCCCCAAGTGGAATCCTGTCAATAGA 758
Db       886 ATCTGGTTTGCCTTATCCTGGATTTGGATCAGTTTCCCAAGTGGTTTCCTGTGAACCGT 945

Qy       759 GAAACGTATATCGAAAGGCTGTCGGCAAGGTATGAAAGAGAGGGTGAGCCTTCTCAGCTT 818
Db       946 GAAACCTACCTCGACAGGCTTGCTTTAAGATATGATCGTGAAGGTGAGCCATCACAGTTA 1005

Qy       819 GCTGGTGTGGATTTTTTCGTGAGTACTGTTGATCCGCTGAAGGAACCGCCATTGATCACT 878
Db      1006 GCTGCTGTTGACATTTTCGTGAGTA TGTTGATCCCTTGAA GAGCCACCCTTGTGACA 1065

Qy       879 GCCAATACAGTCCTTTCCATCTTGCTTGGACTATCCGT GATAAAGTCTCCTGCTAC 938
Db      1066 GCCAACACAGTGCTCTCTATTCTGGCTGTTGACTACCCAGTTGACAAGGTGTCCTGTTAT 1125

Qy       939 GTGTCTGATGATGGTGCAGCTATGCTTTCATTTGAATCTCTGTAGAAACAGCTGAGTTT 998
Db      1126 GTTTCTGATGATGGTGCTGCTATGTTATCATTTGAATCACTTGCAGAAACATCAGAGTTT 1185

Qy       999 GCAAGGAAGTGGGTTCCGTTCTGCAAAAAATTCTCAATTGAACCAAGAGCACCGGAGTTT 1058
Db      1186 GCTCGTAAATGGGTACCATTTTGTAAGAAATATAGCATAGAGCCTCGTGCACCAGAATGG 1245

Qy      1059 TACTTCTCACAGAAAATTGATTACTTGAAAGACAAGGTTCAACCTTCTTTCGTGAAAGAA 1118
Db      1246 TACTTTGCTGCGAAAATAGATTACTTGAAGGATAAAGTTCAGACATCATTTGTCAAAGAT 1305

Qy      1119 CGTAGAGCAATGAAAAGGGATTATGAAGAGTACAAAGTCCGAGTTAATGCCCTGGTAGCA 1178
Db      1306 CGTAGAGCTATGAAGAGGGAATATGAGGAATTTAAAATCCGAATCAATGCACTTGTTTCC 1365

Qy      1179 AAGGCTCAGAAAACACCTGAAGAAGGATGGACTATGCAAGATGGAACACCTTGGCCTGGG 1238
Db      1366 AAAGCCCTAAAATGTCCTGAAGAAGGGTGGGTTATGCAAGATGGCACACCGTGGCCTGGA 1425

Qy      1239 AATAACACACGTGATCACCCTGGGCATGATTCAGGTCTTCCTTGGGAAATACTGGGAGCT 1298
Db      1426 AATAATACAAGGGACCATCC-AGGAATGATCCAGGTCTTCTTAGGGTAAAA- TGGTGGA 1482

Qy      1299 CGTGACATTGAAGGAAATGAACTACCTCGTCTAGTATATGTGTCCAGGGAGAAGAGACCT 1358

```

```
Db      1483 CTTGATGCAGAGGGCAATGAGCTCCCGCGTTTGGTATATGTTTCTCGAGAAAAGCGACCA 1542

Qy      1359 GGCTACCAGCACCACAAAAAGGCTGGTGCAGAAAATGCTCTGGTGAGAGTGTCTGCAGTA 1418
Db      1543 GGATTCCAGCACCACAAAAAGGCTGGTGCTATGAATGCACTGGTGAGAGTTTCAGCAGTT 1602

Qy      1419 CTCACAAATGCTCCCTACATCCTCAATGTTGATTGTGATCACTATGTAAACAATAGCAAG 1478
Db      1603 CTTACCAATGGACCTTTCATCTTGAATCTTGATTGTGATCATTACATAAATAACAGCAAA 1662

Qy      1479 GCTGTTCGAGAGGCAATGTGCATCCTGATGGACCCACAAGTAGGTCGAGATGTATGCTAT 1538
Db      1663 GCCTTAAGAGAAGCAATGTGCTTCCTGATGGACCCAAACCTCGGGAAGCAAGTTTGTTAT 1722

Qy      1539 GTGCAGTTCCCTCAGAGGTTTGATGGCATAGATAAGAGTGATCGCTACGCCAATCGTAAC 1598
Db      1723 GTTCAGTTCCCACAAAGATTTGATGGTATCGATAAGAACGATAGATATGCTAATCGTAAT 1782

Qy      1599 GTAGTTTTCTTTGATGTTAACATGAAAGGGTTGGATGGCATTCAAGGACCAGTATACGTA 1658
Db      1783 ACCGTGTTCTTTGATATTAACTTGAGAGGTTTAGATGGGATTCAAGGACCTGTATATGTC 1842

Qy      1659 GGAACTGGTTGTGTTTTCAACAGGCAAGCACTTTACGGCTACGGGCCTCCTTCTATGCCC 1718
Db      1843 GGAACTGGATGTGTTTTCAACAGAACAGCATTATACGGTTATGAACCTC---CAATAAAA 1899

Qy      1719 AGCTTACGCAAGAGAAAGGATTCTTCATCCTGCTTCTCATGTTGCTGCCCCTCAAAGAAG 1778
Db      1900 GTAAAACACAAGAAGCCAAGTCTTTTATCTAAGCTCTGTGGTGGATCAAGAAAGAAGAAT 1959

Qy      1779 AAGCCTGCTCAAGATCCAGCTGAGGTATACAGAGATGCAAAAAGAGAGGATCTCAATGCT 1838
Db      1960 TCCAAAGCTAAGAAAGAGTCGGACAAAAAGAAATCAGGCAGGCATACTGACTCAACTGTT 2019

Qy      1839 GCCATATTTAATCTTACAGAGATTGATAA---------------TTATGACGAGCATGAA 1883
Db      2020 CCTGTATTCAACCTCGATGACATAGAAGAGGGAGTTGAAGGTGCTGGTTTTGATGATGAA 2079

Qy      1884 AGGTCAATGCTGATCTCCCAGTTGAGCTTTGAGAAAACTTTTGGCTTATCTTCTGTCTTC 1943
Db      2080 AAGGCGCTCTTAATGTCGCAAATGAGCCTGGAGAAGCGATTTGGACAGTCTGCTGTTTTT 2139

Qy      1944 ATTGAGTCTACACTAATGGAGAATGGAGGAGTACCCGAGTCTGCCAACTCACCACCATTC 2003
Db      2140 GTTGCTTCTACCCTAATGGAAAATGGTGGTGTTCCTCCTTCAGCAACTCCAGAAAACCTT 2199

Qy      2004 ATCAAGGAAGCGATTCAAGTCATCGGCTGTGGCTATGAAGAGAAGACTGAATGGGGAAAA 2063
Db      2200 CTCAAAGAGGCTATCCATGTCATTAGTTGTGGTTATGAGGATAAGTCAGATTGGGGAATG 2259

Qy      2064 CAGATTGGTTGGATATATGGGTCAGTCACTGAGGATATCTTAAGTGGCTTCAAGATGCAC 2123
Db      2260 GAGATTGGATGGATCTATGGTTCTGTGACAGAAGATATTCTGACTGGGTTCAAAATGCAT 2319

Qy      2124 TGCCGAGGATGGAGATCAATTTACTGCATGCCCGTAAGGCCTGCATTCAAAGGATCTGCA 2183
Db      2320 GCCCGTGGATGGCGATCCATTTACTGCATGCCTAAGCTTCCAGCTTTCAAGGGTTCTGCT 2379

Qy      2184 CCCATCAACCTGTCTGATAGATTGCACCAGGTCCTCCGATGGGCTCTTGGTTCTGTGGAA 2243
Db      2380 CCTATCAATCTTTCAGATCGTCTGAACCAAGTGCTGAGGTGGGCTTTAGGTTCAGTTGAG 2439

Qy      2244 ATTTTCTTTAGCAGACACTGTCCCCTCTGGTACGGGTTTGGAGGAGGCCGTCTTAAATGG 2303
Db      2440 ATTCTCTTCAGTCGGCATTGTCCTATATGGTATGG---TTACAATGGGAGGCTAAAATTT 2496

Qy      2304 CTCCAAAGGCTTGCGTATATAAACACCATTGTGTACCCATTTACATCCCTCCCTCTCATT 2363
Db      2497 CTTGAGAGGTTTGCGTATGTGAACACCACCATCTACCCTATCACCTCCATTCCTCTTCTC 2556

Qy      2364 GCCTATTGCACAATTCCTGCAGTTTGTCTGCTCACCGGAAAATTCATCATACCAACGCTC 2423
```

```
      2557 ATGTATTGTACATTGCCAGCCGTTTGTCTCTTCACCAACCAGTTTATTATTCCTCAGATT 2616

      2424 TCAAACCTGGCAAGCATGCTGTTTCTTGGCCTCTTTATCTCCATCATTGTAACTGCGGTG 2483
           ||| | ||||| ||   ||||||    ||||| |||||||  | |  || |   |
      2617 AGTAACATTGCAAGTATATGGTTTCTGTCTCTCTTTCTCTCCATTTTCGCCACGGGTATA 2676

      2484 CTTGAGCTAAGATGGAGCGGTGTCAGCATTGAAGATTTATGGCGTAATGAACAATTCTGG 2543
           || ||  | || |||| || ||  |||| || || |  || | | || || || |||
      2677 CTAGAAATGAGGTGGAGTGGCGTAGGCATAGACGAATGGTGGAGAAACGAGCAGTTTTGG 2736

      2544 GTGATCGGAGGTGTTTCAGCCCATCTCTTTGCGGTCTTCCAGGGATTCTTAAAAATGTTG 2603
           || || || || || || || ||| | || || || || || ||  || | ||| |  |
      2737 GTCATTGGTGGAGTATCCGCTCATTTATTCGCTGTGTTTCAAGGTATCCTCAAAGTCCTT 2796

      2604 GCTGGCATCGATACGAACTTCACTGTCACAGCAAAAG---CAGCCGAAGATGCAGAATTT 2660
           || || || || || |||||||| || ||  ||||||   |||  |||| | ||| |||
      2797 GCCGGTATTGACACAAACTTCACAGTTACCTCAAAAGCTTCAGATGAAGACGGAGACTTT 2856

      2661 GGGGAGCTATATATGGTCAAGTGGACAACACTTTTGATTCCTCCAACCACACTTCTCATT 2720
           |  ||||| ||   || |||| ||||||||||||| ||||||| ||||| || || ||||||
      2857 GCTGAGCTCTACTTGTTCAAATGGACAACACTTCTGATTCCGCCAACGACGCTGCTCATT 2916

      2721 ATCAATATGTCGGG---TTGTGCTGGATTCTCTGATGCACTCAACAAAGGATATGAAGCA 2777
           | ||  |  |||   |  ||| ||| ||||| |||| ||||||  ||||  || ||
      2917 GTAAACTTAGTGGGAGTTGTTGCAGGAGTCTCTTATGCTATCAACAGTGGATACCAATCA 2976

      2778 TGGGGGCCTCTCTTTGGCAAGGTGTTCTTTGCTTTCTGGGTGATTCTTCATCTCTATCCA 2837
           ||||| || |||||||| ||| |||||||||| |||||||||||  ||||  | || ||
>     2977 TGGGGACCACTCTTTGGTAAGTTGTTCTTTGCCTTCTGGGTGATTGTTCACTTGTACCCT 3036

/     2838 TTCCTTAAAGGTCTAATGGGTCGCCAAAACCTAACACCAACCATTGTTGTTCTCTGGTCA 2897
           ||||| || ||| | |||||||| || |||  || || |||||||||||  |||||||
>     3037 TTCCTCAAGGGTTTGATGGGTCGACAGAACCGGACTCCTACCATTGTTGTGGTCTGGTCT 3096

/     2898 GTGCTGTTGGCCTCTGTCTTCTCTCTCGTTTGGGTCAAGATCAATCCATTCGTTAACAAA 2957
           || || ||||| ||| |||||||  |  | |||||  | |||  |||| |||  ||  |  |
>     3097 GTTCTCTTGGCTTCTATCTTCTCGTTGTTGTGGGTTAGGATTGATCCCTTCACTAGCCGA 3156

/     2958 GTTGATAACACCTTGGTTGCGGA 2980
           ||  |  | |    ||  |||
>     3157 GTCACTGGCCCGGACATTCTGGA 3179


ESULT 15
AV06567
D   AAV06567 standard; cDNA; 3614 BP.
X
C   AAV06567;
X
T   06-JUL-1998  (first entry)
X
E   Arabidopsis cellulose biosynthetic gene clone Ath-B.
X
W   Cellulose; cellulose synthase; RSW1 gene; beta-1,4-glucan;
W   transgenic plant; ss.
X
S   Arabidopsis thaliana var. Columbia.
X
H   Key             Location/Qualifiers
T   CDS             217..3411
T                   /*tag= a
X
N   WO9800549-A1.
X
D   08-JAN-1998.
X
F   24-JUN-1997;   97WO-AU00402.
X
R   27-JUN-1996;   96AU-0000699.
X
A   (CSIR ) COMMONWEALTH SCI & IND RES ORG.
A   (AUSU ) UNIV AUSTRALIAN NAT.
X
```

```
PI   Arioli T, Bonzon AS, Peng L, Williamson RE;
XX
DR   WPI; 1998-086974/08.
DR   P-PSDB; AAW33819.
XX
PT   DNA encoding cellulose biosynthetic enzyme useful for manipulation
PT   of cellulose and beta 1,4 glucan
XX
PS   Claim 12; Page 138-143; 200pp; English.
XX
CC   cDNA clone Ath-B was isolated from an Arabidopsis thaliana cDNA
CC   library using probes obtained by PCR with primers (see AAT95632-34)
CC   based on cellulose synthase RSW1 genomic clone 23H12 (see AAV06563)
CC   and EST clone AAT20782 (see AAV06562). It is closely related to RSW1
CC   cDNA (see AAV06565) and to partial genomic clone 12C4 (see AAV06564),
CC   and encodes a 1065 polypeptide (see AAW33819). Claimed nucleic acid
CC   molecules (see AAV06562-69) coding for claimed polypeptides (see
CC   AAW33816-20 and AAW46202) involved in cellulose biosynthesis can be
CC   used to manipulate the cellulose and/or beta glucan content of
CC   transgenic plants. Expression of certain genes in the sense
CC   orientation increases the level of cellulose and reduces the level
CC   of non crystalline beta-1,4-glucans in plants, providing plants
CC   with modified primary cell wall properties, e.g. fibre properties, or
CC   having increased resistance to stresses or pests. Antisense,
CC   ribozyme or co-suppression molecules can be used to reduce the
CC   cellulose content of a transgenic plant, e.g. to improve
CC   digestibility or to alter carbon partitioning such that increased
CC   carbon is available for growth, rather than deposited as cellulose.
XX
SQ   Sequence 3614 BP; 963 A; 717 C; 963 G; 1971 T; 0 other;

  Query Match             35.1%;  Score 1133.8;  DB 19;  Length 3614;
  Best Local Similarity   68.1%;  Pred. No. 3.6e-290;
  Matches 1691;  Conservative    0;  Mismatches  792;  Indels   30;  Gaps    7;

Qy        519 CCGTCTGCAAGAGGCTTCGGAGCCGCTTTCAATTGTTTATCCAATTCCACCGCAACAAGCTC 578
              | |  ||  | ||  | ,||  |  ||.          | |||'|||          | |
Db        922 CTGAATGACGAAGCGAGGCAGCTTCTGTCAACGAAAGTTTCAATTCCTTCATCACGGATC 981

Qy        579 ACACCATACAGAGCAGTGATCATTATGCGACTGGTCATTCTGGGCCTCTTCTTCCACTTC 638
              |  ||  |||.||   || ||  |  .||| |  || |  ||  | ,||||.|  || | |
Db        982 AATCCTTACAGAATGGTTATTATGCTGCGCTTGTTATCCTTTGTCTCTTCTTCATTAC 1041

Qy        639 AGAATAACAAATCCTGTCGATAGTGCCTTTGGTCTGTGGCTTACTTCTGTCATATGTGAG 698
              | |||:||| || ||    | |||||, |  |  |||||    ||:|, ||||:|||
Db       1042 CGTATAACAAACCCAGTGCCAAATGCCTTTGCTCTATGGCTGGTCTCTGTGATATGTGAG 1101

Qy        699 ATCTGGTTTGCATTTTCTTGGGTGTTGGATCAGTTCCCCAAGTGGAATCCTGTCAATAGA 758
              |||||||'.| |: || ||: | ||||||.'|| ||||||||  ||||| |,  |
Db       1102 ATCTGGTTTGCCTTATCCTGGATTTTGGATCAGTTTCCCAAGTGGTTTCCTGTGAACCGT 1161

Qy        759 GAAACGTATATCGAAAGGCTGTCGGCAAGGTATGAAAGAGAGGGTGAGCCTTCTCAGCTT 818
              ||||| ||  |||| ||||  |  '|| ||'|  | || ||||'|| || ||| :
Db       1162 GAAACCTACCTCGACAGGCTTGCTTTAAGATATGATCGTGAAGGTGAGCCATCACAGTTA 1221

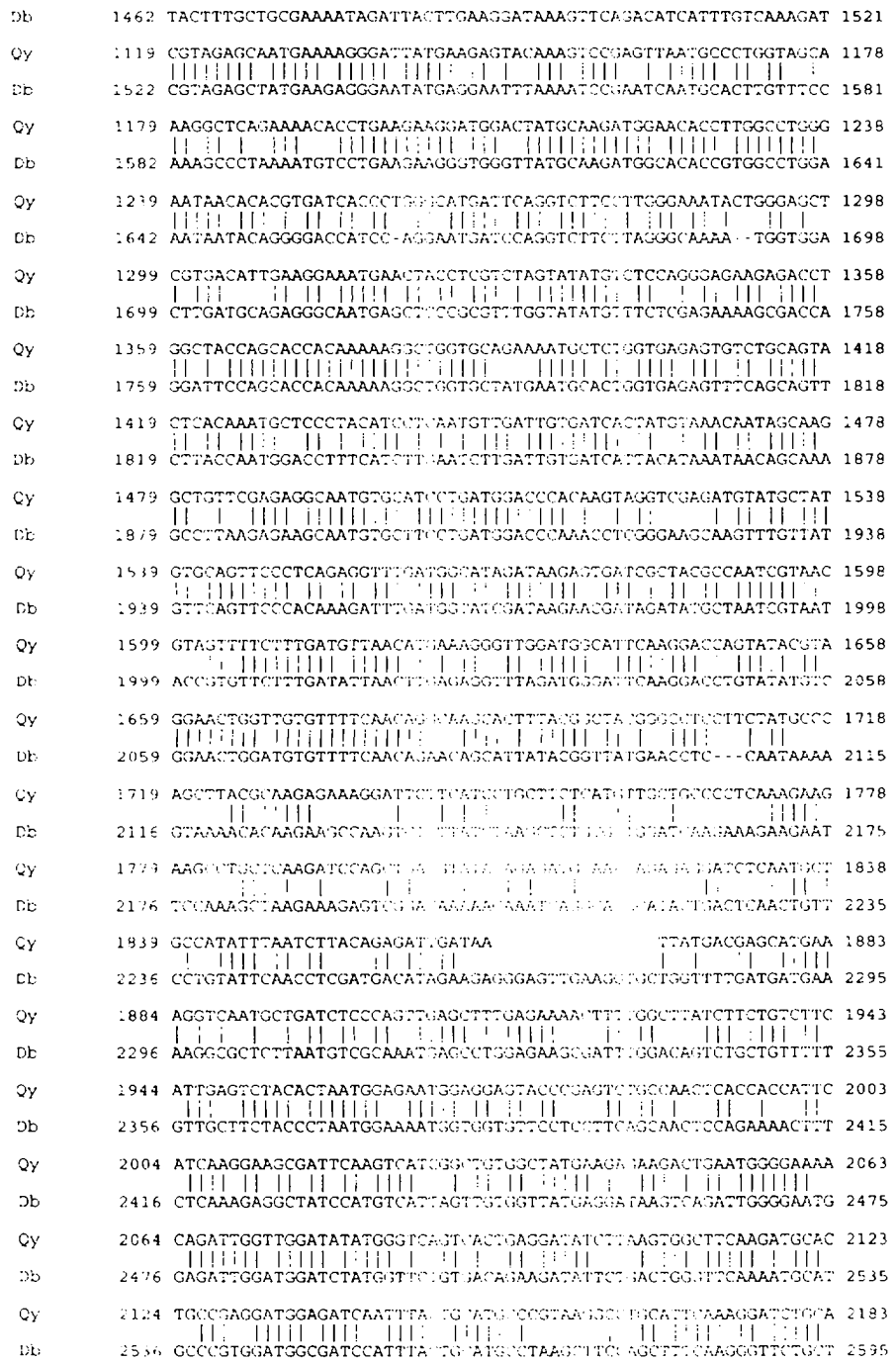
Qy        819 GCTGGTGTGGATTTTTTCGTGAGTACTGTTGATCCGCTGAAGGAACCGCCATTGATCACT 878
              |||| ||' ||  ||||||||||||||||'|| ::  |||||||  |  ||  :  | |.
Db       1222 GCTGCTGTTGACATTTTCGTGAGTACTGTTGA'CCCTTGAAGGAGCCACCCCTTGTGACA 1281

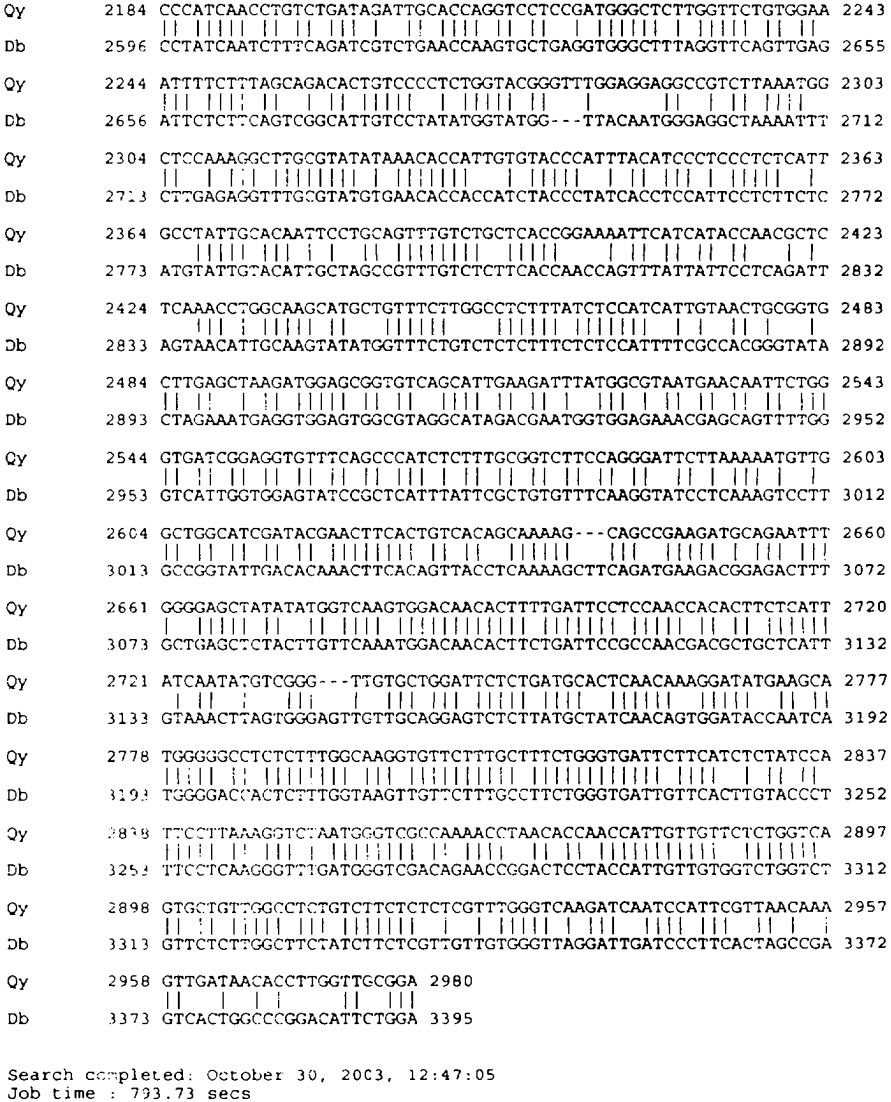
Qy        879 GCCAATACAGTCCTTTCCATCCTTGCTGTGGA'TATCCCGTCGATAAAGTCTCCTGCTAC 938
              ||||| ||||| |, ||  ||  ||'|  || :, || || ||  ||  ||| :|
Db       1282 GCCAACACAGTGCTCTCTATTCTGGCTGTTGA'TACCCAGTTGACAAGGTGTCCTGTTAT 1341

Qy        939 GTGTCTGATGATGGTGCAGCTATGCTTTCATTTGAATCTCTTGTAGAAACAGCTGAGTTT 998
              || |.|||| |||||| ,|||| | || ||| .'|| .||| |.|||  . |||.|
Db       1342 GTTTCTGATGATGGTGCTGCTATGTTATCATTTGAATCACTTGCAGAAACATCAGAGTTT 1401

Qy        999 GCAAGGAAGTGGGTTCCGTTCTGCAAAAAATT'TCAATTGAACCAAGAGCACCGGAGTTT 1058
              ||  | || |||| || || |||,| |||     :  || ||  | |'|| |' |
Db       1402 GCTCGTAAATGGGTACCATTTTGCAAGAAATATAGCATAGAGCCTCGTGCACCAGAATGG 1461

Qy       1059 TACTTCTCACAGAAAATTGATTACTTGAAAGACAAGGTTCAACCTTCTTTCGTGAAAGAA 1118
              |||||  :  |||||| |||'|||||| || |: |||||  | || || || ||||:
```





Search completed: October 30, 2003, 12:47:05
Job time : 793.73 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 30, 2003, 05:38:04 ; Search time 170.411 Seconds
(without alignments)
8371.232 Million cell updates/sec

Title: US-09-980-043-1
Perfect score: 3232
Sequence: 1 gtcgacccacgcgtccgtct.....aagaaaaaaaaaaaaaaaaa 3232

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1989.8	61.6	3328	3	US-08-960-048-1	Sequence 1, Appli
2	1989.8	61.6	3328	4	US-09-838-586-1	Sequence 1, Appli
3	1133.8	35.1	3614	4	US-09-221-013A-9	Sequence 9, Appli
4	1097.8	34.0	3603	4	US-09-221-013A-5	Sequence 5, Appli
5	1096.2	33.9	3673	4	US-09-221-013A-11	Sequence 11, Appl
6	1043	32.3	3828	4	US-09-221-013A-7	Sequence 7, Appli
7	775.4	24.0	2248	4	US-09-221-013A-1	Sequence 1, Appli
8	446.8	13.8	1741	4	US-09-221-013A-13	Sequence 13, Appl
9	352.8	10.9	8411	4	US-09-221-013A-3	Sequence 3, Appli
10	194.6	6.0	5009	4	US-09-221-013A-4	Sequence 4, Appli
11	138.2	4.3	291	4	US-09-313-294A-5519	Sequence 5519, Ap
12	136	4.2	281	4	US-09-313-294A-6485	Sequence 6485, Ap
13	135	4.2	268	4	US-09-313-294A-849	Sequence 849, App
14	129.8	4.0	300	4	US-09-313-294A-4753	Sequence 4753, Ap
15	127	3.9	283	4	US-09-313-294A-1981	Sequence 1981, Ap
16	117.2	3.6	286	4	US-09-313-294A-4696	Sequence 4696, Ap
17	102.8	3.2	287	4	US-09-313-294A-5274	Sequence 5274, Ap
18	94.4	2.9	250	4	US-09-313-294A-582	Sequence 582, App
19	93.8	2.9	278	4	US-09-313-294A-4015	Sequence 4015, Ap
20	80.4	2.5	284	4	US-09-313-294A-3993	Sequence 3993, Ap
21	74.2	2.3	317	4	US-09-313-294A-7123	Sequence 7123, Ap
22	63	1.9	159	4	US-09-313-294A-1693	Sequence 1693, Ap
23	56.2	1.7	4612	3	US-08-960-048-2	Sequence 2, Appli
24	56.2	1.7	4612	4	US-09-838-586-2	Sequence 2, Appli
C 25	56	1.7	7218	1	US-08-232-463-14	Sequence 14, Appl
26	48.8	1.5	2920	3	US-09-276-400-1	Sequence 1, Appli
27	48.8	1.5	2920	4	US-09-448-076-1	Sequence 1, Appli
28	48.8	1.5	2920	4	US-09-702-572-1	Sequence 1, Appli
29	45.2	1.4	1063	3	US-08-960-048-3	Sequence 3, Appli
30	45.2	1.4	1063	4	US-09-838-586-3	Sequence 3, Appli
31	45	1.4	591	4	US-09-276-599-7	Sequence 7, Appli
32	44.4	1.4	7218	1	US-08-232-463-14	Sequence 14, Appl
33	44	1.4	1024	4	US-09-328-475C-74	Sequence 74, Appl
34	43.4	1.3	732	4	US-09-149-476-66	Sequence 66, Appl
35	42.8	1.3	1788	1	US-08-035-928-1	Sequence 1, Appli
36	42.6	1.3	1355	4	US-09-370-838-31	Sequence 31, Appl
37	42.6	1.3	1835	4	US-09-564-808-5	Sequence 5, Appli
38	42.4	1.3	1443	4	US-09-904-615-21	Sequence 21, Appl
39	42.2	1.3	2294	2	US-08-967-508-10	Sequence 10, Appl
40	42.2	1.3	2294	3	US-08-967-506-10	Sequence 10, Appl
41	42.2	1.3	2294	5	PCT-US94-02552-10	Sequence 10, Appl
42	41.6	1.3	1230	3	US-09-387-574-5	Sequence 5, Appli
43	41.6	1.3	1230	4	US-09-668-096-5	Sequence 5, Appli
44	41.6	1.3	1265	3	US-08-991-789A-169	Sequence 169, App
45	41.6	1.3	1265	4	US-09-062-451-169	Sequence 169, App

ALIGNMENTS

RESULT 1
US-08-960-048-1
; Sequence 1, Application US/08960048C
; Patent No. 6271443
; GENERAL INFORMATION:
; APPLICANT: Stalker, D. et al.
; TITLE OF INVENTION: Plant Cellulose Synthase and Promoter
; TITLE OF INVENTION: Sequences
; FILE REFERENCE: 15621/01/US
; CURRENT APPLICATION NUMBER: US/08/960,048C
; CURRENT FILING DATE: 1997-10-29
; PRIOR APPLICATION NUMBER: 60/029,987
; PRIOR FILING DATE: 1996-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 3328
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Oligonucleotide
US-08-960-048-1

Query Match 61.6%; Score 1989.8; DB 3; Length 3328;
Best Local Similarity 80.8%; Pred. No. 0;
Matches 2399; Conservative 0; Mismatches 552; Indels 18; Gaps 6;

```
Qy          40 AAAGAGCTGGTAAAGTGGTAATAAGCAAGATGATGGAATCTGGGGCTCCTATATGCCATA 99
               || |||  |  ||||     |    | ||||||||||||||||| |||| | ||||| |
Db         150 AAGGAGAAAGATAAGTACTTTTTTTGAGAATGATGGAATCTGGGGTTCCTGTTTGCCACA 209

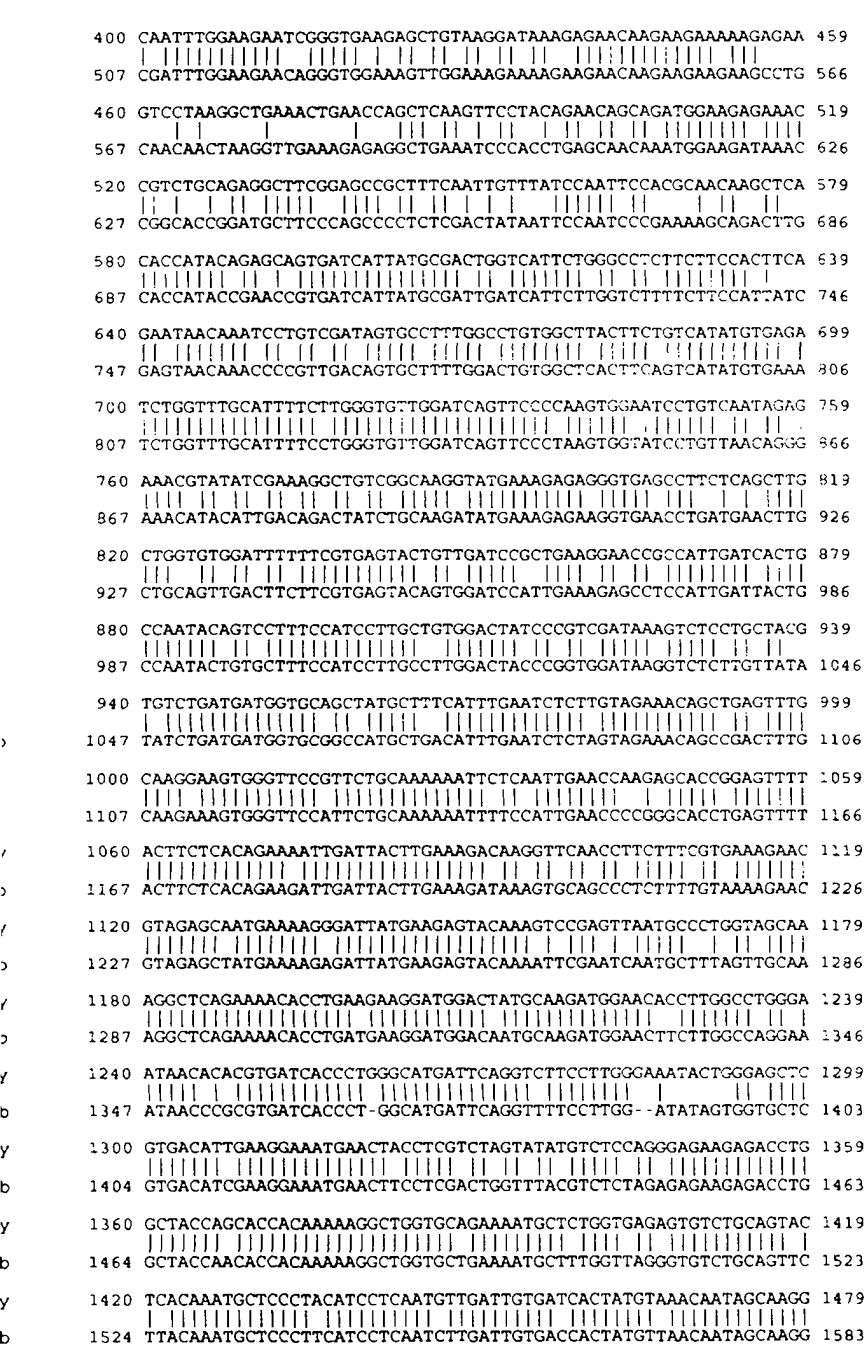
Qy         100 CCTGTGGTGAACAGGTGGGGCATGATGCAAATGGGGAGCTATTTGTGGCTTGCCATGAGT 159
               | |||||||||||  || |||    |||  |||| || |  |||||||||||||||| |
Db         210 CTTGTGGTGAACATGTTGGGTTGAATGTTAATGGTGAACCTTTTGTGGCTTGCCATGAAT 269

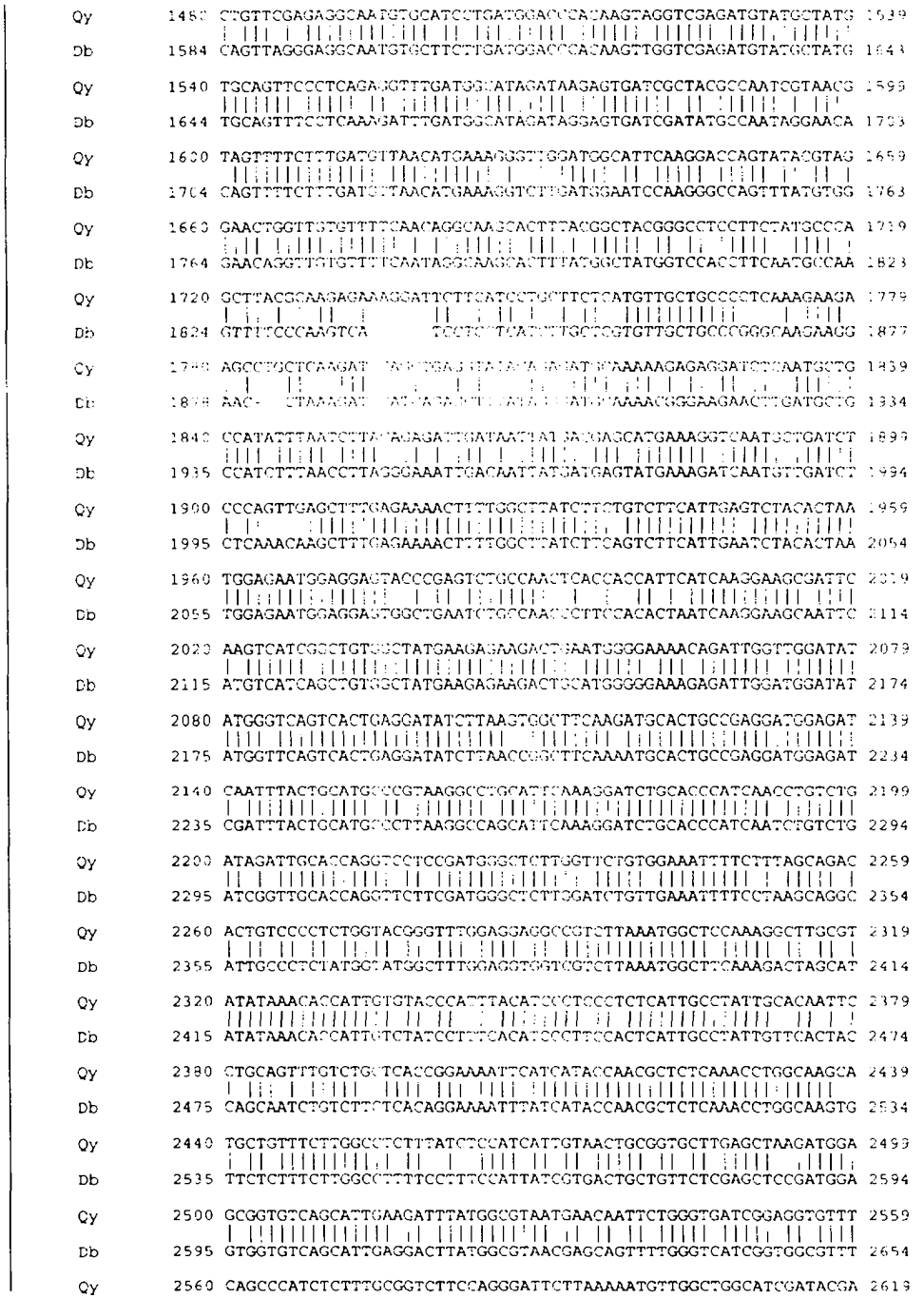
Qy         160 GTAGCTATCCCATGTGCAAGTCTTGTTTCGAGTTTGAAATCAATGAGGGCCGGAAAGTTT 219
               |||  |  || || || |||  |||||| |||| |||  | || || || || |||| ||
Db         270 GTAATTTCCCTATTTGTAAGAGTTGTTTTGAGTATGATCTTAAGGAAGGACGAAAAGCTT 329

Qy         220 GCTTGCGGTGTGGCTCGCCATATGATGAGAACTTGCTGGATGATGTAGAAAAGAAGGGGT 279
               ||||||| |||||    |||||||||| ||| || |||| ||||| ||   |||||
Db         330 GCTTGCGTTGTGGTAGTCCATATGATGAAAACCTGTTGGACGATGTCGA---GAAGGCCA 386

Qy         280 CTGGCAATCAATCCACAATGGCATCTCACCTCAACGATTCTCAGGATGTCGGAATCCATG 339
               | ||| ||||||| ||||||||  | ||  | ||| | |||||||||| ||||| ||||
Db         387 CCGGCGATCAATCGACAATGGCTGCACATTTGAACAAGTCTCAGGATGTTGGAATTCATG 446

Qy         340 CTAGACATATCAGTAGTGTGTCCACTGTGGATAGTGAAATGAATGATGAATATGGGAATC 399
               | |||||||||| |||||||| ||  |||||||||||||  ||| ||  ||||||||
Db         447 CAAGACATATCAGCAGTGTGTCTACATTGGATAGTGAAATGGCTGAAGACAATGGGAATT 506
```


400 CAATTTGGAAGAATCGGGTGAAGAGCTGTAAGGATAAAGAGAACAAGAAGAAAAAGAGAA 459
507 CGATTTGGAAGAACAGGGTGGAAAGTTGGAAAGAAAAGAAGAACAAGAAGAAGAAGCCTG 566
460 GTCCTAAGGCTGAAACTGAACCAGCTCAAGTTCCTACAGAACAGCAGATGGAAGAGAAAC 519
567 CAACAACTAAGGTTGAAAGAGAGGCTGAAATCCCACCTGAGCAACAAATGGAAGATAAAC 626
520 CGTCTGCAGAGGCTTCGGAGCCGCTTTCAATTGTTTATCCAATTCCACGCAACAAGCTCA 579
627 CGGCACCGGATGCTTCCCAGCCCCTCTCGACTATAATTCCAATCCCGAAAAGCAGACTTG 686
580 CACCATACAGAGCAGTGATCATTATGCGACTGGTCATTCTGGGCCTCTTCTTCCACTTCA 639
687 CACCATACCGAACCGTGATCATTATGCGATTGATCATTCTTGGTCTTTTCTTCCATTATC 746
640 GAATAACAAATCCTGTCGATAGTGCCTTTGGCCTGTGGCTTACTTCTGTCATATGTGAGA 699
747 GAGTAACAAACCCCGTTGACAGTGCTTTTGGACTGTGGCTCACTTCAGTCATATGTGAAA 806
700 TCTGGTTTGCATTTTCTTGGGTGTTGGATCAGTTCCCCAAGTGGAATCCTGTCAATAGAG 759
807 TCTGGTTTGCATTTTCCTGGGTGTTGGATCAGTTCCCTAAGTGGTATCCTGTTAACAGGG 866
760 AAACGTATATCGAAAGGCTGTCGGCAAGGTATGAAAGAGAGGGTGAGCCTTCTCAGCTTG 819
867 AAACATACATTGACAGACTATCTGCAAGATATGAAAGAGAAGGTGAACCTGATGAACTTG 926
820 CTGGTGTGGATTTTTTCGTGAGTACTGTTGATCCGCTGAAGGAACCGCCATTGATCACTG 879
927 CTGCAGTTGACTTCTTCGTGAGTACAGTGGATCCATTGAAAGAGCCTCCATTGATTACTG 986
880 CCAATACAGTCCTTTCCATCCTTGCTGTGGACTATCCCGTCGATAAAGTCTCCTGCTACG 939
987 CCAATACTGTGCTTTCCATCCTTGCCTTGGACTACCCGGTGGATAAGGTCTCTTGTTATA 1046
940 TGTCTGATGATGGTGCAGCTATGCTTTCATTTGAATCTCTTGTAGAAACAGCTGAGTTTG 999
1047 TATCTGATGATGGTGCGGCCATGCTGACATTTGAATCTCTAGTAGAAACAGCCGACTTTG 1106
1000 CAAGGAAGTGGGTTCCGTTCTGCAAAAAATTCTCAATTGAACCAAGAGCACCGGAGTTTT 1059
1107 CAAGAAAGTGGGTTCCATTCTGCAAAAAATTTTCCATTGAACCCCGGGCACCTGAGTTTT 1166
1060 ACTTCTCACAGAAAATTGATTACTTGAAAGACAAGGTTCAACCTTCTTTCGTGAAAGAAC 1119
1167 ACTTCTCACAGAAGATTGATTACTTGAAAGATAAAGTGCAGCCCTCTTTTGTAAAAGAAC 1226
1120 GTAGAGCAATGAAAAGGGATTATGAAGAGTACAAAGTCCGAGTTAATGCCCTGGTAGCAA 1179
1227 GTAGAGCTATGAAAAGAGATTATGAAGAGTACAAAATTCGAATCAATGCTTTAGTTGCAA 1286
1180 AGGCTCAGAAAACACCTGAAGAAGGATGGACTATGCAAGATGGAACACCTTGGCCTGGGA 1239
1287 AGGCTCAGAAAACACCTGATGAAGGATGGACAATGCAAGATGGAACTTCTTGGCCAGGAA 1346
1240 ATAACACACGTGATCACCCTGGGCATGATTCAGGTCTTCCTTGGGAAATACTGGGAGCTC 1299
1347 ATAACCCGCGTGATCACCCT-GGCATGATTCAGGTTTTCCTTGG--ATATAGTGGTGCTC 1403
1300 GTGACATTGAAGGAAATGAACTACCTCGTCTAGTATATGTCTCCAGGGAGAAGAGACCTG 1359
1404 GTGACATCGAAGGAAATGAACTTCCTCGACTGGTTTACGTCTCTAGAGAGAAGAGACCTG 1463
1360 GCTACCAGCACCACAAAAAGGCTGGTGCAGAAAATGCTCTGGTGAGAGTGTCTGCAGTAC 1419
1464 GCTACCAACACCACAAAAAGGCTGGTGCTGAAAATGCTTTGGTTAGGGTGTCTGCAGTTC 1523
1420 TCACAAATGCTCCCTACATCCTCAATGTTGATTGTGATCACTATGTAAACAATAGCAAGG 1479
1524 TTACAAATGCTCCCTTCATCCTCAATCTTGATTGTGACCACTATGTTAACAATAGCAAGG 1583


Qy 1480 CTGTTCGAGAGGCAATGTGCATCCTGATGGACCCACAAGTAGGTCGAGATGTATGCTATG 1539
Db 1584 CAGTTAGGGAGGCAATGTGCTTCTTGATGGACCCACAAGTTGGTCGAGATGTATGCTATG 1643
Qy 1540 TGCAGTTCCCTCAGAAGTTTGATGGCATAGATAAGAGTGATCGCTACGCCAATCGTAACG 1599
Db 1644 TGCAGTTTCCTCAAAGATTTGATGGCATAGATAGGAGTGATCGATATGCCAATAGGAACA 1703
Qy 1600 TAGTTTTCTTTGATGTTAACATGAAAGGGTTGGATGGCATTCAAGGACCAGTATACGTAG 1659
Db 1704 CAGTTTTCTTTGATGTTAACATGAAAGGTCTTGATGGAATCCAAGGGCCAGTTTATGTGG 1763
Qy 1660 GAACTGGTTGTGTTTTCAACAGGCAAGCACTTTACGGCTACGGGCCTCCTTCTATGCCCA 1719
Db 1764 GAACAGGTTGTGTTTTCAATAGGCAAGCACTTTATGGCTATGGTCCACCTTCAATGCCAA 1823
Qy 1720 GCTTACGCAAGAGAAAGGATTCTTCATCCTGCTTCTCATGTTGCTGCCCCTCAAAGAAGA 1779
Db 1824 GTTTTCCCAAGTCA TCCTCTTCATCTTTGCTCGTGTTGCTGCCCGGGCAAGAAGG 1877
Qy 1780 AGCCTGCTCAAGAT [illegible] AAAAAGAGAGGATCTCAATGCTG 1839
Db 1878 AAC- CTAAAGAT [illegible] AAAACGGGAAGAACTTGATGCTG 1934
Qy 1840 CCATATTTAATCTTATAGAGATTGATAATAATAGGAGCATGAAAGGTCAATGCTGATCT 1899
Db 1935 CCATCTTTAACCTTAGGGAAATTGACAATTATGATGAGTATGAAAGATCAATGTTGATCT 1994
Qy 1900 CCCAGTTGAGCTTTGAGAAAACTTTTGGCTTATCTTCTGTCTTCATTGAGTCTACACTAA 1959
Db 1995 CTCAAACAAGCTTTGAGAAAACTTTTGGCTTATCTTCAGTCTTCATTGAATCTACACTAA 2054
Qy 1960 TGGAGAATGGAGGAGTACCCGAGTCTGCCAACTCACCACCATTCATCAAGGAAGCGATTC 2019
Db 2055 TGGAGAATGGAGGAGTGGCTGAATCTGCCAACTCTTCCACACTAATCAAGGAAGCAATTC 2114
Qy 2020 AAGTCATCGGCTGTGGCTATGAAGAGAAGACTGCATGGGGGAAAAGAGATTGGATGGATAT 2079
Db 2115 ATGTCATCAGCTGTGGCTATGAAGAGAAGACTGCATGGGGGAAAAGAGATTGGATGGATAT 2174
Qy 2080 ATGGGTCAGTCACTGAGGATATCTTAAGTGGCTTCAAGATGCACTGCCGAGGATGGAGAT 2139
Db 2175 ATGGTTCAGTCACTGAGGATATCTTAACCGGCTTCAAAATGCACTGCCGAGGATGGAGAT 2234
Qy 2140 CAATTTACTGCATGCCCGTAAGGCCTGCATTCAAAGGATCTGCACCCATCAACCTGTCTG 2199
Db 2235 CGATTTACTGCATGCCCTTAAGGCCAGCATTCAAAGGATCTGCACCCATCAATCTGTCTG 2294
Qy 2200 ATAGATTGCACCAGGTCCTCCGATGGGCTCTTGGTTCTGTGGAAATTTTCTTTAGCAGAC 2259
Db 2295 ATCGGTTGCACCAGGTTCTTCGATGGGCTCTTGGATCTGTTGAAATTTTCCTAAGCAGGC 2354
Qy 2260 ACTGTCCCCTCTGGTACGGGTTTGGAGGAGGCCGTCTTAAATGGCTCCAAAGGCTTGCGT 2319
Db 2355 ATTGCCCTCTATGGCATGGCTTTGGAGGTGGTCGTCTTAAATGGCTTCAAAGACTAGCAT 2414
Qy 2320 ATATAAACACCATTGTGTACCCATTTACATCCCTCCCTCTCATTGCCTATTGCACAATTC 2379
Db 2415 ATATAAACACCATTGTCTATCCTTTCACATCCCTTCCACTCATTGCCTATTGTTCACTAC 2474
Qy 2380 CTGCAGTTTGTCTGCTCACCGGAAAATTCATCATACCAACGCTCTCAAACCTGGCAAGCA 2439
Db 2475 CAGCAATCTGTCTTCTCACAGGAAAATTTATCATACCAACGCTCTCAAACCTGGCAAGTG 2534
Qy 2440 TGCTGTTTCTTGGCCTCTTTATCTCCATCATTGTAACTGCGGTGCTTGAGCTAAGATGGA 2499
Db 2535 TTCTCTTTCTTGGCCTTTTCCTTTCCATTATCGTGACTGCTGTTCTCGAGCTCCGATGGA 2594
Qy 2500 GCGGTGTCAGCATTGAAGATTTATGGCGTAATGAACAATTCTGGGTGATCGGAGGTGTTT 2559
Db 2595 GTGGTGTCAGCATTGAGGACTTATGGCGTAACGAGCAGTTTTGGGTCATCGGTGGCGTTT 2654
Qy 2560 CAGCCCATCTCTTTGCGGTCTTCCAGGGATTCTTAAAAATGTTGGCTGGCATCGATACGA 2619

```
Db       2655 CAGCCCATCTCTTTGCCGTCTTCCAAGGTTTCCTTAAGATGCTTGCGGCATTGACACCA 2714

Qy       2620 ACTTCACTGTCACAGCAAAAGCAGCCGAAGATGCAGAATTTGGGGAGCTATATATGGTCA 2679

Db       2715 ACTTTACTGTCACTGCCAAAGCAGCTGATGATGCAGATTTTGGTGAGCTCTACATTGTGA 2774

Qy       2680 AGTGGACAACACTTTTGATTCCTCCAACCACACTTCTCATTATCAATATGTCGGGT   T 2736

Db       2775 AATGGACTACACTTCTAATCCCTCCAACAACACTCCTCATCGTCAACATGGTTGGTGTCG 2834

Qy       2737 GTGCTGGATTCTCTGATGCAGTCAACAAAGGATATGAAGCATGGGGGCCTCTCTTTGGCA 2796

Db       2835 TTGCCGGATTCTCCGATGCCCTCAACAAAGGGTACGAAGCTTGGGGACCACTCTTTGGCA 2894

Qy       2797 AGGTGTTCTTTGCTTTCTGGGTGATTCTTCATCTCTATCCATTCCTTAAAGGTCTAATGG 2856

Db       2895 AAGTGTTCTTTTCCTTCTGGGTCATCCTCCATCTTTATCCATTCCTCAAAGGTCTTATGG 2954

Qy       2857 GTCGCCAAAACCTAACACCAACCATTGTTGTTCTCTGGTCAGTGCTGTTGGCCTCTGTCT 2916

Db       2955 GACGCCAAAACAGGACACCAACCATTGTTGTCCTTTGGTCAGTGTTGTTGGCTTCTGTCT 3014

Qy       2917 TCTCTCTCGTTTGGGTCAAGATCAATCCATTCGTTAACAAATTGATAACACCTTGGTTG 2976

Db       3015 TCTCTCTTGTTTGGGTTCGGATCAACCCGTTTGTCAGCACCGCCGATAGCACCACCGTGT 3074

Qy       2977 CGGAGACCTGCATTTCCATTGATTGCTGA 3005

Db       3075 CACAGAGCTGCATTTCCATTGATTGTTGA 3103


RESULT 2
US-09-838-586-1
; Sequence 1, Application US/09838586
; Patent No. 6576818
; GENERAL INFORMATION:
;  APPLICANT: Stalker, D. et al.
;  TITLE OF INVENTION: Plant Cellulose Synthase and Promoter
;  TITLE OF INVENTION:  Sequences
;  FILE REFERENCE: 15621/02/US
;  CURRENT APPLICATION NUMBER: US/09/838,586
;  CURRENT FILING DATE:  2001-04-18
;  PRIOR APPLICATION NUMBER: 60/029,943
;  PRIOR FILING DATE: 1996-10-29
;  PRIOR APPLICATION NUMBER: 08/[illegible]
;  PRIOR FILING DATE: 1997-10-29
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: PastSEQ for Windows Version 4.0
; SEQ ID NO 1
;  LENGTH: 3328
;  TYPE: DNA
;  ORGANISM: Artificial Sequence
;  FEATURE:
;  OTHER INFORMATION: Synthetic Oligonucleotide
US-09-838-586-1

  Query Match             61.6%;  Score 1989.8;  DB 4;  Length 3328;
  Best Local Similarity   80.8%;  Pred. No. 0;
  Matches 2399;  Conservative    0;  Mismatches  552;  Indels   18;  Gaps    6;

Qy         40 AAAGAGCTGGTAAAGTGGTAAAAGCAAGATGATGGAATCTGGGCTCCTATATGCCATA 99

Db        150 AAGGAGAAAGATAAGTACTTTTTTTGAGAATGATGGAATCTGGGTTCCTGTTTGCCACA 209

Qy        100 CCTGTGGTGAACAGGTGGGGCATGATTCAAATGGGGAGCTAATTGTGGCTTGCCATGAGT 159

Db        210 CTTGTGGTGAACATGTTGGGTTGAATTTAATGATGAACCTTTGTGGCTTGCCATGAAT 269

Qy        160 GTAGCTATCCCATGTGCAAGTTTGGTTCTGAGTTTGAAATCATGAGGGCCGGAAAGTTT 219

Db        270 GTAATTTCCCTATTTGTAAGATTGTTTGAGTATGATCTTAGGAATTACGAAAAGCTT 329

Qy        220 GCTTGCGGTGTGGCTCGCCATATGATGAGAACTTGCTGGATGATGTAGAAAAGAAGGGGT 279

Db        330 GCTTGCGTTGTGGTAGTCCATATGATGAAAACCTGTTGGACGATGTCGA---GAAGGCCA 386

Qy        280 CTGGCAATCAATCCACAATGGCATCTCACCTCAACGATTCTCAGGATGTCGGAATCCATG 339

Db        387 CCGGCGATCAATCGACAATGGCTGCACATTTGAACAAGTCTCAGGATGTTGGAATTCATG 446

Qy        340 CTAGACATATCAGTAGTGTGTCCACTGTGGATAGTGAAATGAATGATGAATATGGGAATC 399

Db        447 CAAGACATATCAGCAGTGTGTCTACATTGGATAGTGAAATGGCTGAAGACAATGGGAATT 506

Qy        400 CAATTTGGAAGAATCGGGTGAAGAGCTGTAAGGATAAAGAGAACAAGAAGAAAAAGAGAA 459

Db        507 CGATTTGGAAGAACAGGGTGGAAAGTTGGAAAGAAAAGAAGAACAAGAAGAAGAAGCCTG 566

Qy        460 GTCCTAAGGCTGAAACTGAACCAGCTCAAGTTCCTACAGAACAGCAGATGGAAGAGAAAC 519

Db        567 CAACAACTAAGGTTGAAAGAGAGGCTGAAATCCCACCTGAGCAACAAATGGAAGATAAAC 626

Qy        520 CGTCTGCAGAGGCTTCGGAGCCGCTTTCAATTGTTTATCCAATTCCACGCAACAAGCTCA 579

Db        627 CGGCACCGGATGCTTCCCAGCCCCTCTCGACTATAATTCCAATCCCGAAAAGCAGACTTG 686

Qy        580 CACCATACAGAGCAGTGATCATTATGCGACTGGTCATTCTGGGCCTCTTCTTCCACTTCA 639

Db        687 CACCATACCGAACCGTGATCATTATGCGATTGATCATTCTTGGTCTTTTCTTCCATTATC 746

Qy        640 GAATAACAAATCCTGTCGATAGTGCCTTTGGCCTGTGGCTTACTTCTGTCATATGTGAGA 699

Db        747 GAGTAACAAACCCCGTTGACAGTGCTTTTGGACTGTGGCTCACTTCAGTCATATGTGAAA 806

Qy        700 TCTGGTTTGCATTTTCTTGGGTGTTGGATCAGTTCCCCAAGTGGAATCCTGTCAATAGAG 759

Db        807 TCTGGTTTGCATTTTCCTGGGTGTTGGATCAGTTCCCTAAGTGGTATCCTGTTAACAGGG 866

Qy        760 AAACGTATATCGAAAGGCTGTCGGCAAGGTATGAAAGAGAGGGTGAGCCTTCTCAGCTTG 819

Db        867 AAACATACATTGACAGACTATCTGCAAGATATGAAAGAGAAGGTGAACCTGATGAACTTG 926

Qy        820 CTGGTGTGGATTTTTTCGTGAGTACTGTTGATCCGCTGAAGGAACCGCCATTGATCACTG 879

Db        927 CTGCAGTTGACTTCTTCGTGAGTACAGTGGATCCATTGAAAGAGCCTCCATTGATTACTG 986

Qy        880 CCAATACAGTCCTTTCCATCCTTGCTGTGGACTATCCCGTCGATAAAGTCTCCTGCTACG 939

Db        987 CCAATACTGTGCTTTCCATCCTTGCCTTGGACTACCCGGTGGATAAGGTCTCTTGTTATA 1046

Qy        940 TGTCTGATGATGGTGCAGCTATGCTTTCATTTGAATCTCTTGTAGAAACAGCTGAGTTTG 999

Db       1047 TATCTGATGATGGTGCGGCCATGCTGACATTTGAATCTCTAGTAGAAACAGCCGACTTTG 1106

Qy       1000 CAAGGAAGTGGGTTCCGTTCTGCAAAAAATTCTCAATTGAACCAAGAGCACCGGAGTTTT 1059

Db       1107 CAAGAAAGTGGGTTCCATTCTGCAAAAAATTTTCCATTGAACCCCGGGCACCTGAGTTTT 1166

Qy       1060 ACTTCTCACAGAAAATTGATTACTTGAAAGACAAGGTTCAACCTTCTTTCGTGAAAGAAC 1119

Db       1167 ACTTCTCACAGAAGATTGATTACTTGAAAGATAAAGTGCAGCCCTCTTTTGTAAAAGAAC 1226

Qy       1120 GTAGAGCAATGAAAAGGGATTATGAAGAGTACAAAGTCCGAGTTAATGCCCTGGTAGCAA 1179

Db       1227 GTAGAGCTATGAAAAGAGATTATGAAGAGTACAAAATTCGAATCAATGCTTTAGTTGCAA 1286

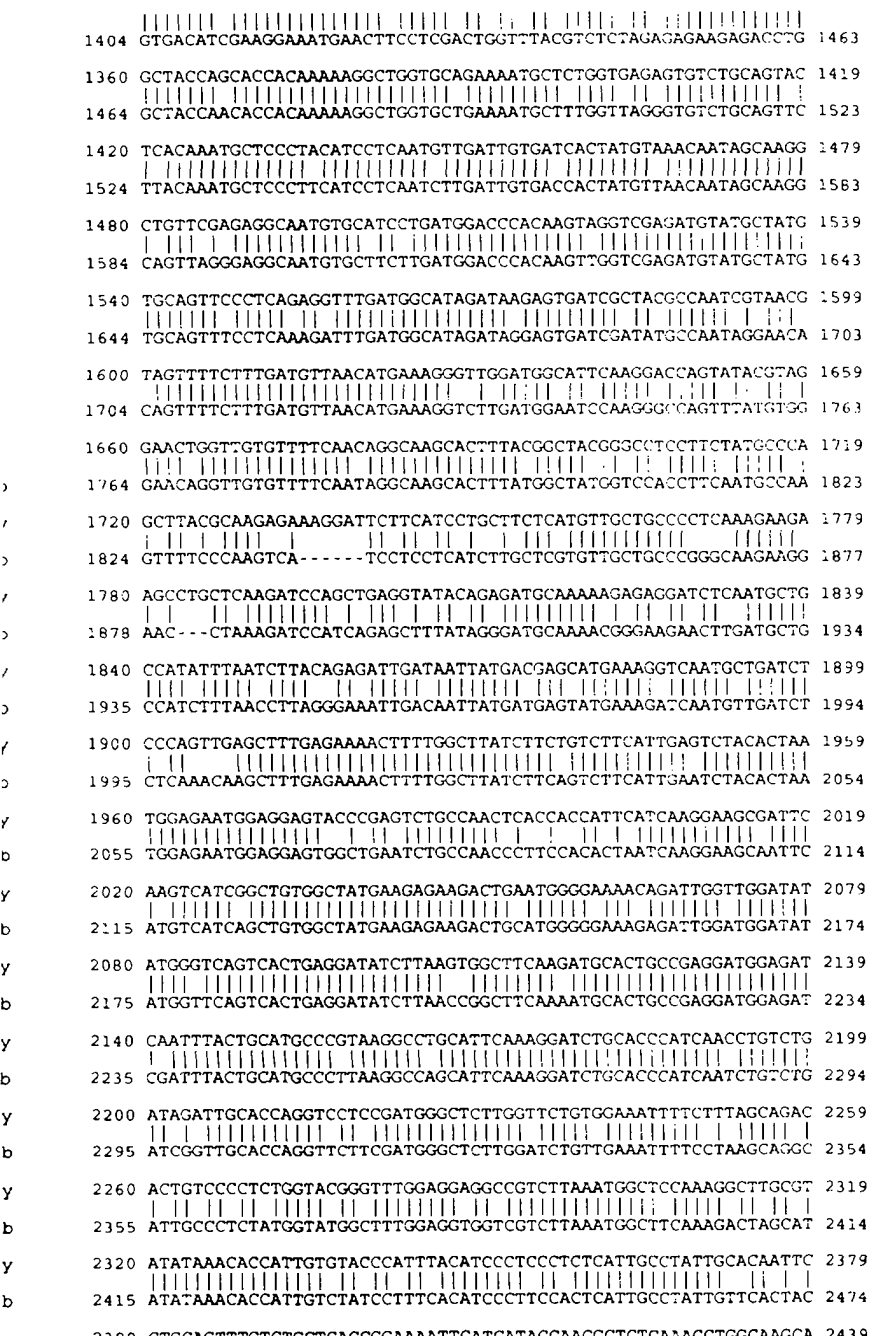
Qy       1180 AGGCTCAGAAAACACCTGAAGAAGGATGGACTATGCAAGATGGAACACCTTGGCCTGGGA 1239

Db       1287 AGGCTCAGAAAACACCTGATGAAGGATGGACAATGCAAGATGGAACTTCTTGGCCAGGAA 1346

Qy       1240 ATAACACACGTGATCACCCTGGGCATGATTCAGGTCTTCCTTGGGAAATACTGGGAGCTC 1299

Db       1347 ATAACCCGCGTGATGACCCT-GGCATGATTCAGGTTTTCCTTGG--ATATAGTGGTGCTC 1403

Qy       1300 GTGACATTGAAGGAAATGAACTACCTCGTCTAGTATATGTCTCCAGGGAGAAGAGACCTG 1359
```



```
1404 GTGACATCGAAGGAAATGAACTTCCTCGACTGGTTTACGTCTCTAGAGAGAAGACCTG 1463

1360 GCTACCAGCACCACAAAAAGGCTGGTGCAGAAAATGCTCTGGTGAGAGTGTCTGCAGTAC 1419
1464 GCTACCAACACCACAAAAAGGCTGGTGCTGAAAATGCTTTGGTTAGGGTGTCTGCAGTTC 1523

1420 TCACAAATGCTCCCTACATCCTCAATGTTGATTGTGATCACTATGTAAACAATAGCAAGG 1479
1524 TTACAAATGCTCCCTTCATCCTCAATCTTGATTGTGACCACTATGTTAACAATAGCAAGG 1583

1480 CTGTTCGAGAGGCAATGTGCATCCTGATGGACCCACAAGTAGGTCGAGATGTATGCTATG 1539
1584 CAGTTAGGGAGGCAATGTGCTTCTTGATGGACCCACAAGTTGGTCGAGATGTATGCTATG 1643

1540 TGCAGTTCCCTCAGAGGTTTGATGGCATAGATAAGAGTGATCGCTACGCCAATCGTAACG 1599
1644 TGCAGTTTCCTCAAAGATTTGATGGCATAGATAGGAGTGATCGATATGCCAATAGGAACA 1703

1600 TAGTTTTCTTTGATGTTAACATGAAAGGGTTGGATGGCATTCAAGGACCAGTATACGTAG 1659
1704 CAGTTTTCTTTGATGTTAACATGAAAGGTCTTGATGGAATCCAAGGGCCAGTTTATGTGG 1763

1660 GAACTGGTTGTGTTTTCAACAGGCAAGCACTTTACGGCTACGGGCCTCCTTCTATGCCCA 1719
1764 GAACAGGTTGTGTTTTCAATAGGCAAGCACTTTATGGCTATGGTCCACCTTCAATGCCAA 1823

1720 GCTTACGCAAGAGAAAGGATTCTTCATCCTGCTTCTCATGTTGCTGCCCCTCAAAGAAGA 1779
1824 GTTTTCCCAAGTCA------TCCTCCTCATCTTGCTCGTGTTGCTGCCCGGGCAAGAAGG 1877

1780 AGCCTGCTCAAGATCCAGCTGAGGTATACAGAGATGCAAAAAGAGAGGATCTCAATGCTG 1839
1878 AAC---CTAAAGATCCATCAGAGCTTTATAGGGATGCAAAACGGGAAGAACTTGATGCTG 1934

1840 CCATATTTAATCTTACAGAGATTGATAATTATGACGAGCATGAAAGGTCAATGCTGATCT 1899
1935 CCATCTTTAACCTTAGGGAAATTGACAATTATGATGAGTATGAAAGATCAATGTTGATCT 1994

1900 CCCAGTTGAGCTTTGAGAAAACTTTTGGCTTATCTTCTGTCTTCATTGAGTCTACACTAA 1959
1995 CTCAAACAAGCTTTGAGAAAACTTTTGGCTTATCTTCAGTCTTCATTGAATCTACACTAA 2054

1960 TGGAGAATGGAGGAGTACCCGAGTCTGCCAACTCACCACCATTCATCAAGGAAGCGATTC 2019
2055 TGGAGAATGGAGGAGTGGCTGAATCTGCCAACCCTTCCACACTAATCAAGGAAGCAATTC 2114

2020 AAGTCATCGGCTGTGGCTATGAAGAGAAGACTGAATGGGGAAAACAGATTGGTTGGATAT 2079
2115 ATGTCATCAGCTGTGGCTATGAAGAGAAGACTGCATGGGGGAAAGAGATTGGATGGATAT 2174

2080 ATGGGTCAGTCACTGAGGATATCTTAAGTGGCTTCAAGATGCACTGCCGAGGATGGAGAT 2139
2175 ATGGTTCAGTCACTGAGGATATCTTAACCGGCTTCAAAATGCACTGCCGAGGATGGAGAT 2234

2140 CAATTTACTGCATGCCCGTAAGGCCTGCATTCAAAGGATCTGCACCCATCAACCTGTCTG 2199
2235 CGATTTACTGCATGCCCTTAAGGCCAGCATTCAAAGGATCTGCACCCATCAATCTGTCTG 2294

2200 ATAGATTGCACCAGGTCCTCCGATGGGCTCTTGGTTCTGTGGAAATTTTCTTTAGCAGAC 2259
2295 ATCGGTTGCACCAGGTTCTTCGATGGGCTCTTGGATCTGTTGAAATTTTCCTAAGCAGGC 2354

2260 ACTGTCCCCTCTGGTACGGGTTTGGAGGAGGCCGTCTTAAATGGCTCCAAAGGCTTGCGT 2319
2355 ATTGCCCTCTATGGTATGGCTTTGGAGGTGGTCGTCTTAAATGGCTTCAAAGACTAGCAT 2414

2320 ATATAAACACCATTGTGTACCCATTTACATCCCTCCCTCTCATTGCCTATTGCACAATTC 2379
2415 ATATAAACACCATTGTCTATCCTTTCACATCCCTTCCACTCATTGCCTATTGTTCACTAC 2474

2380 CTGCAGTTTGTCTGCTCACCGGAAAATTCATCATACCAACGCTCTCAAACCTGGCAAGCA 2439
```

```
Db       2435 CAGCAATCTGTCTT TGACAGGAAAATTTATCATACCAACGCTCTCAAACCTGGCAAGTG 2494

Qy       2440 TGCTGTTTCTTGGC TCTTTATCTCTATCATTGTAACTGCGGTGCTTGAGCTAAGATGGA 2499
Db       2495 TTCTCTTTCTTGGC TTTGCTTTCCATTATCGTGACTGCTGTTCTCGAGCTCCGATGGA 2554

Qy       2500 GCGGTGTCAGCATC AAGATTTATGCGTAATGAACAATTCTGGGTGATCGGAGGTGTTT 2559
Db       2555 GTGGTGTCAGCATT AGGACTTATGGCGTAACGAGCAGTTTTGGGTCATCGGTGGCGTTT 2614

Qy       2560 CAGCCCATCTCTTT TGGTCTTCCAGGATTCTTAAAAATGTTGGCTGGCATCGATACGA 2619
Db       2615 CAGCCCATCTCTTT TCGTCTTCCAAGGTTTCGTTAAGATGCTTGCGGGCATTGACACCA 2674

Qy       2620 ACTTCACTGTCACA AAAAGCAGCCGAAGATGCAGAATTTGGGGAGCTATATATGGTCA 2679
Db       2675 ACTTTACTGTCACT AAAGCAGCTGATGATGCAGATTTTGGTGAGCTCTACATTGTGA 2734

Qy       2680 AGTGGACAAGACTT TGACTTCTTCAACTACATTTCTCATTATCAATATGTCGGGT---T 2736
Db       2735 AATGGACTATACTT AAGTCTTCAACTACATTTCTCATCGTCAACATGGTTGGTGTCG 2794

Qy       2737 GTGGTGGATTCTCT AAAAAAGAAGGAAGCATGGGGGCCTCTCTTTGGCA 2786
Db       2795 TTGGCGGATTCTCC ATGCCTTAACAAA AAAGAAGCTTGGGGACCACTCTTTGGCA 2844

Qy       2787 AGGTGTTCTTTGCT TCTGGGTGATTCTT ATCTCTATCCATTCCTTAAAGGTCTAATGG 2846
Db       2845 AAGTGTTCTTTTCC TCTGGGTCATCTC TATCCATTCCTCAAAGGTCTTATGG 2894

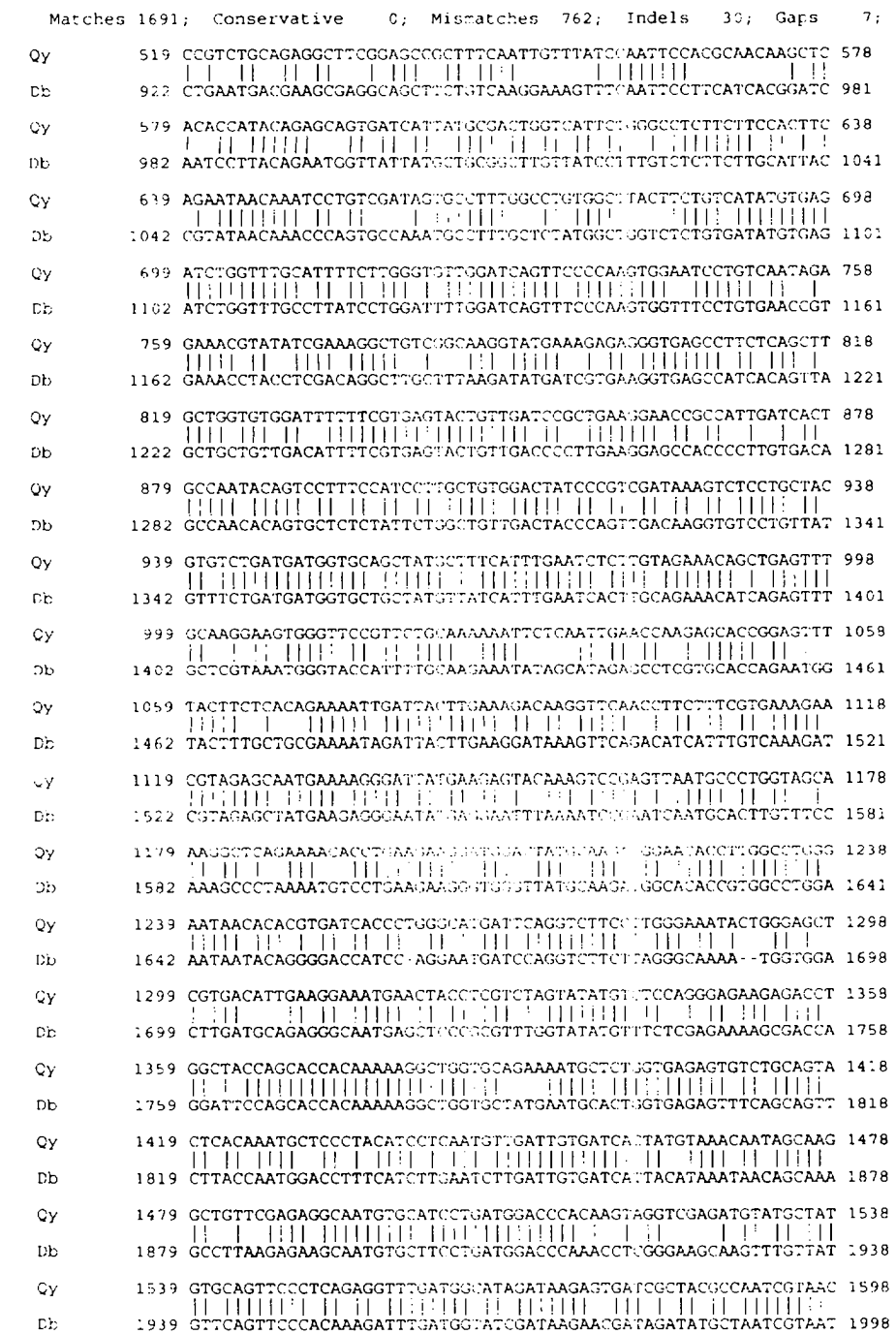
Qy       2847 GTCGCCAAAACCTAACACCAACCATTGTTGTT TCTGGTCAGTGCTGTTGGCCTCTGTCT 2906
Db       2895 GACGCCAAAACAGGATACCAACCATTGTTGTC TTTGGTCAGTGTTGTTGGCTTCTGTCT 2954

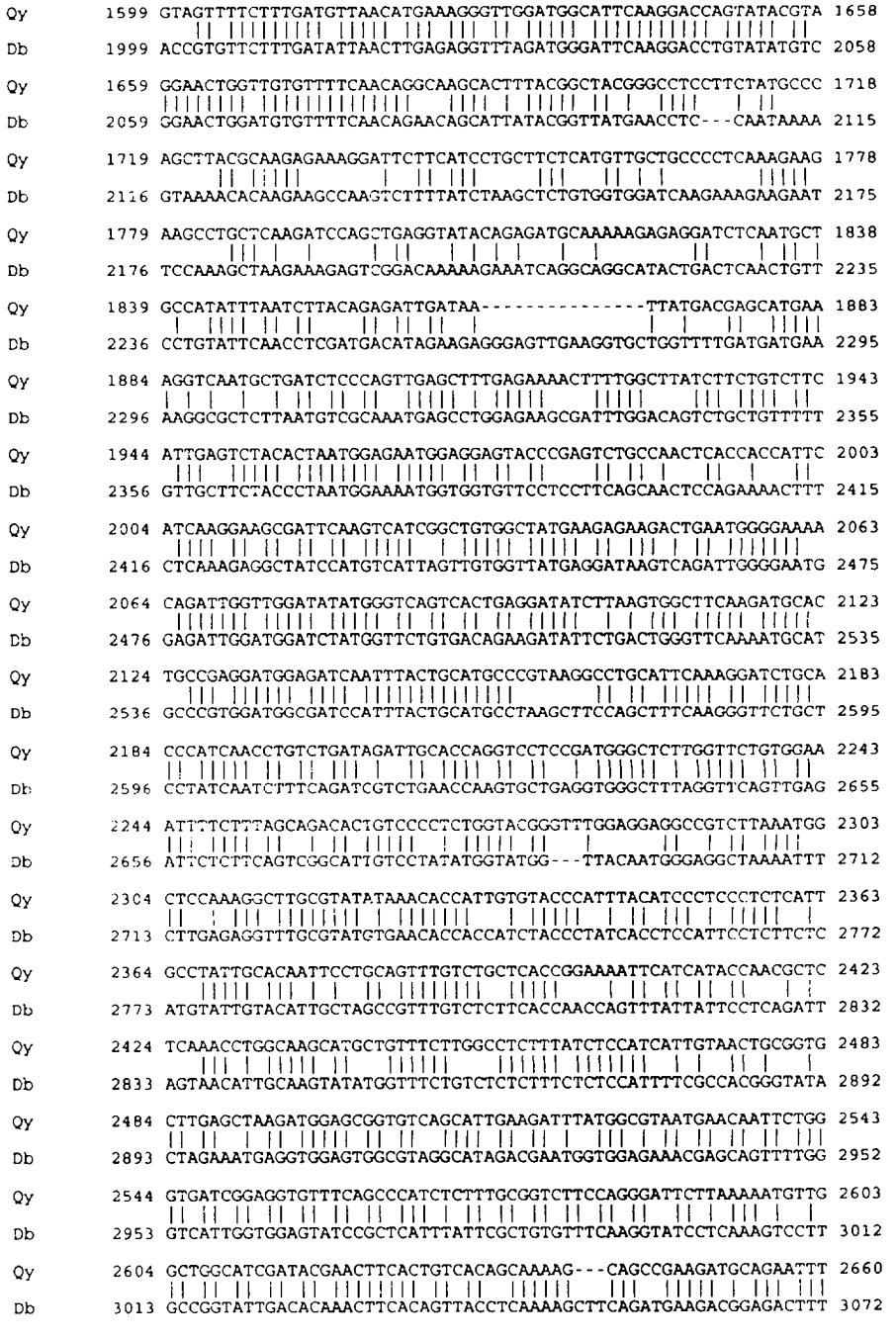
Qy       2907 TCTCTCTCTGTTTGG TCAAGATCAATCCATTC TAACAAAGTTGATAACACCTTGGTTG 2966
Db       2955 TCTCTCTTGTTTGG TTCGGATCAATCCGTTT TAAGCACCGCCGATAGCACCACCGTGT 3014

Qy       2967 CGGAGACCTGCATT CATTGATTGTTGA 3014
Db       3015 CACAGAGCTGCATT CATTGATTGTTGA 3043
```

```
RESULT 3
US-09-221-013A 9
; Sequence 9, Application US/09221013A
; Patent No. 6495740
; GENERAL INFORMATION:
;  APPLICANT:  Arioli, Antonio
;  APPLICANT:  Williamson, Richard E.
;  APPLICANT:  Betzner, Andreas S.
;  APPLICANT:  Peng, Liangcai
;  TITLE OF INVENTION:  Manipulation of cellulose and/or Beta-1,4-glucan
;  FILE REFERENCE:  96-98
;  CURRENT APPLICATION NUMBER:  US/09/221,013A
;  CURRENT FILING DATE:  1998-12-23
;  PRIOR APPLICATION NUMBER:  PCT/AU97/00405
;  PRIOR FILING DATE:  1997-6-24
;  PRIOR APPLICATION NUMBER:  AU PO0699
;  PRIOR FILING DATE:  1996-6-27
;  NUMBER OF SEQ ID NOS:  37
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 9
;  LENGTH: 3614
;  TYPE: DNA
;  ORGANISM: Arabidopsis thaliana
;  FEATURE:
;  NAME/KEY: CDS
;  LOCATION: (217)..(3411)
US-09-221-013A 9

  Query Match           35.1%;  Score 1133.8;  DB 4;  Length 3614;
  Best Local Similarity  68.1%;  Pred. No. 8.3e-305;
```


Matches 1691; Conservative 0; Mismatches 762; Indels 38; Gaps 7;
Qy 519 CCGTCTGCAGAGGCTTCGGAGCCGCTTTCAATTGTTTATGCAATTCCACGCAACAAGCTC 578
Db 922 CTGAATGACGAAGCGAGGCAGCTTCTGTCAAGGAAAGTTTCAATTCCTTCATCACGGATC 981
Qy 579 ACACCATACAGAGCAGTGATCATTATGCGACTGGTCATTCTGGGCCTCTTCTTCCACTTC 638
Db 982 AATCCTTACAGAATGGTTATTATGCTGCGGCTTGTTATCCTTTGTCTCTTCTTGCATTAC 1041
Qy 639 AGAATAACAAATCCTGTCGATAGTGCCTTTGGCCTGTGGCTTACTTCTGTCATATGTGAG 698
Db 1042 CGTATAACAAACCCAGTGCCAAATGCCTTTGCTCTATGGCTGGTCTCTGTGATATGTGAG 1101
Qy 699 ATCTGGTTTGCATTTTCTTGGGTGTTGGATCAGTTCCCCAAGTGGAATCCTGTCAATAGA 758
Db 1102 ATCTGGTTTGCCTTATCCTGGATTTTGGATCAGTTTCCCAAGTGGTTTCCTGTGAACCGT 1161
Qy 759 GAAACGTATATCGAAAGGCTGTCGGCAAGGTATGAAAGAGAGGGTGAGCCTTCTCAGCTT 818
Db 1162 GAAACCTACCTCGACAGGCTTGCTTTAAGATATGATCGTGAAGGTGAGCCATCACAGTTA 1221
Qy 819 GCTGGTGTGGATTTTTTCGTGAGTACTGTTGATCCGCTGAAGGAACCGCCATTGATCACT 878
Db 1222 GCTGCTGTTGACATTTTCGTGAGTACTGTTGACCCCTTGAAGGAGCCACCCCTTGTGACA 1281
Qy 879 GCCAATACAGTCCTTTCCATCCTTGCTGTGGACTATCCCGTCGATAAAGTCTCCTGCTAC 938
Db 1282 GCCAACACAGTGCTCTCTATTCTGGCTGTTGACTACCCAGTTGACAAGGTGTCCTGTTAT 1341
Qy 939 GTGTCTGATGATGGTGCAGCTATGCTTTCATTTGAATCTCTTGTAGAAACAGCTGAGTTT 998
Db 1342 GTTTCTGATGATGGTGCTGCTATGTTATCATTTGAATCACTTGCAGAAACATCAGAGTTT 1401
Qy 999 GCAAGGAAGTGGGTTCCGTTCTGCAAAAAATTCTCAATTGAACCAAGAGCACCGGAGTTT 1058
Db 1402 GCTCGTAAATGGGTACCATTTTGCAAGAAATATAGCATAGAGCCTCGTGCACCAGAATGG 1461
Qy 1059 TACTTCTCACAGAAAATTGATTACTTGAAAGACAAGGTTCAACCTTCTTTCGTGAAAGAA 1118
Db 1462 TACTTTGCTGCGAAAATAGATTACTTGAAGGATAAAGTTCAGACATCATTTGTCAAAGAT 1521
Qy 1119 CGTAGAGCAATGAAAAGGGATTATGAAGAGTACAAAGTCCGAGTTAATGCCCTGGTAGCA 1178
Db 1522 CGTAGAGCTATGAAGAGGGAATATGAAAAATTTAAAATCCGAATCAATGCACTTGTTTCC 1581
Qy 1179 AAGGCTCAGAAAACACCTGAAGAAGAAGTATATGCAATTGGAACACCTGGCCTGGG 1238
Db 1582 AAAGCCCTAAAATGTCCTGAAGAAGGTGTTATGCAAGAGGCACACCGTGGCCTGGA 1641
Qy 1239 AATAACACACGTGATCACCCTGGGCATGATTCAGGTCTTCCTTGGGAAATACTGGGAGCT 1298
Db 1642 AATAATACAGGGGACCATCCAGGAATGATCCAGGTCTTCTTAGGGCAAAA--TGGTGGA 1698
Qy 1299 CGTGACATTGAAGGAAATGAACTACCTCGTCTAGTATATGTTCCAGGGAGAAGAGACCT 1358
Db 1699 CTTGATGCAGAGGGCAATGAGCTCCGCGTTTGGTATATGTTTCTCGAGAAAAGCGACCA 1758
Qy 1359 GGCTACCAGCACCACAAAAAGGCTGGTGCAGAAAATGCTCTGGTGAGAGTGTCTGCAGTA 1418
Db 1759 GGATTCCAGCACCACAAAAAGGCTGGTGCTATGAATGCACTGGTGAGAGTTTCAGCAGTT 1818
Qy 1419 CTCACAAATGCTCCCTACATCCTCAATGTTGATTGTGATCACTATGTAAACAATAGCAAG 1478
Db 1819 CTTACCAATGGACCTTTCATCTTGAATCTTGATTGTGATCACTACATAAATAACAGCAAA 1878
Qy 1479 GCTGTTCGAGAGGCAATGTGCATCCTGATGGACCCACAAGTAGGTCGAGATGTATGCTAT 1538
Db 1879 GCCTTAAGAGAAGCAATGTGCTTCCTGATGGACCCAAACCTCGGGAAGCAAGTTTGTTAT 1938
Qy 1539 GTGCAGTTCCCTCAGAGGTTTGATGGCATAGATAAGAGTGATCGCTACGCCAATCGTAAC 1598
Db 1939 GTTCAGTTCCCACAAAGATTTGATGGTATCGATAAGAACGATAGATATGCTAATCGTAAT 1998


Qy 1599 GTAGTTTTCTTTGATGTTAACATGAAAGGGTTGGATGGCATTCAAGGACCAGTATACGTA 1658
Db 1999 ACCGTGTTCTTTGATATTAACTTGAGAGGTTTAGATGGGATTCAAGGACCTGTATATGTC 2058
Qy 1659 GGAACTGGTTGTGTTTTCAACAGGCAAGCACTTTACGGCTACGGGCCTCCTTCTATGCCC 1718
Db 2059 GGAACTGGATGTGTTTTCAACAGAACAGCATTATACGGTTATGAACCTC---CAATAAAA 2115
Qy 1719 AGCTTACGCAAGAGAAAGGATTCTTCATCCTGCTTCTCATGTTGCTGCCCCTCAAAGAAG 1778
Db 2116 GTAAAACACAAGAAGCCAAGTCTTTTATCTAAGCTCTGTGGTGGATCAAGAAAGAAGAAT 2175
Qy 1779 AAGCCTGCTCAAGATCCAGCTGAGGTATACAGAGATGCAAAAAGAGAGGATCTCAATGCT 1838
Db 2176 TCCAAAGCTAAGAAAGAGTCGGACAAAAAGAAATCAGGCAGGCATACTGACTCAACTGTT 2235
Qy 1839 GCCATATTTAATCTTACAGAGATTGATAA----------TTATGACGAGCATGAA 1883
Db 2236 CCTGTATTCAACCTCGATGACATAGAAGAGGGAGTTGAAGGTGCTGGTTTTGATGATGAA 2295
Qy 1884 AGGTCAATGCTGATCTCCCAGTTGAGCTTTGAGAAAACTTTTGGCTTATCTTCTGTCTTC 1943
Db 2296 AAGGCGCTCTTAATGTCGCAAATGAGCCTGGAGAAGCGATTTGGACAGTCTGCTGTTTTT 2355
Qy 1944 ATTGAGTCTACACTAATGGAGAATGGAGGAGTACCCGAGTCTGCCAACTCACCACCATTC 2003
Db 2356 GTTGCTTCTACCCTAATGGAAAATGGTGGTGTTCCTCCTTCAGCAACTCCAGAAAACTTT 2415
Qy 2004 ATCAAGGAAGCGATTCAAGTCATCGGCTGTGGCTATGAAGAGAAGACTGAATGGGGAAAA 2063
Db 2416 CTCAAAGAGGCTATCCATGTCATTAGTTGTGGTTATGAGGATAAGTCAGATTGGGGAATG 2475
Qy 2064 CAGATTGGTTGGATATATGGGTCAGTCACTGAGGATATCTTAAGTGGCTTCAAGATGCAC 2123
Db 2476 GAGATTGGATGGATCTATGGTTCTGTGACAGAAGATATTCTGACTGGGTTCAAAATGCAT 2535
Qy 2124 TGCCGAGGATGGAGATCAATTTACTGCATGCCCGTAAGGCCTGCATTCAAAGGATCTGCA 2183
Db 2536 GCCCGTGGATGGCGATCCATTTACTGCATGCCTAAGCTTCCAGCTTTCAAGGGTTCTGCT 2595
Qy 2184 CCCATCAACCTGTCTGATAGATTGCACCAGGTCCTCCGATGGGCTCTTGGTTCTGTGGAA 2243
Db 2596 CCTATCAATCTTTCAGATCGTCTGAACCAAGTGCTGAGGTGGGCTTTAGGTTCAGTTGAG 2655
Qy 2244 ATTTTCTTTAGCAGACACTGTCCCCTCTGGTACGGGTTTGGAGGAGGCCGTCTTAAATGG 2303
Db 2656 ATTCTCTTCAGTCGGCATTGTCCTATATGGTATGG---TTACAATGGGAGGCTAAAATTT 2712
Qy 2304 CTCCAAAGGCTTGCGTATATAAACACCATTGTGTACCCATTTACATCCCTCCCTCTCATT 2363
Db 2713 CTTGAGAGGTTTGCGTATGTGAACACCACCATCTACCCTATCACCTCCATTCCTCTTCTC 2772
Qy 2364 GCCTATTGCACAATTCCTGCAGTTTGTCTGCTCACCGGAAAATTCATCATACCAACGCTC 2423
Db 2773 ATGTATTGTACATTGCTAGCCGTTTGTCTCTTCACCAACCAGTTTATTATTCCTCAGATT 2832
Qy 2424 TCAAACCTGGCAAGCATGCTGTTTCTTGGCCTCTTTATCTCCATCATTGTAACTGCGGTG 2483
Db 2833 AGTAACATTGCAAGTATATGGTTTCTGTCTCTCTTTCTCTCCATTTTCGCCACGGGTATA 2892
Qy 2484 CTTGAGCTAAGATGGAGCGGTGTCAGCATTGAAGATTTATGGCGTAATGAACAATTCTGG 2543
Db 2893 CTAGAAATGAGGTGGAGTGGCGTAGGCATAGACGAATGGTGGAGAAACGAGCAGTTTTGG 2952
Qy 2544 GTGATCGGAGGTGTTTCAGCCCATCTCTTTGCGGTCTTCCAGGGATTCTTAAAAATGTTG 2603
Db 2953 GTCATTGGTGGAGTATCCGCTCATTTATTCGCTGTGTTTCAAGGTATCCTCAAAGTCCTT 3012
Qy 2604 GCTGGCATCGATACGAACTTCACTGTCACAGCAAAAG---CAGCCGAAGATGCAGAATTT 2660
Db 3013 GCCGGTATTGACACAAACTTCACAGTTACCTCAAAAGCTTCAGATGAAGACGGAGACTTT 3072

```
        2661 GGGGAGCTATATATGGTCAAGTGGACAACACTTTTGATTCCTCCAACCACACTTCTCATT 2720
             |  |||||  ||  || |||| |||||||||||||| |||||||  |||||  ||  || ||||||
        3073 GCTGAGCTCTACTTGTTCAAATGGACAACACTTCTGATTCCGCCAACGACGCTGCTCATT 3132

        2721 ATCAATATGTCGGG---TTGTGCTGGATTCTCTGATGCACTCAACAAAGGATATGAAGCA 2777
             | ||  |   |||   |  ||| ||| ||||| ||||  ||||||  |||||  || ||
        3133 GTAAACTTAGTGGGAGTTGTTGCAGGAGTCTCTTATGCTATCAACAGTGGATACCAATCA 3192

        2778 TGGGGGCCTCTCTTTGGCAAGGTGTTCTTTGCTTTCTGGGTGATTCTTCATCTCTATCCA 2837
             ||||| || |||||||| ||| |||||||||| |||||||||||| ||||  | || ||
        3193 TGGGGACCACTCTTTGGTAAGTTGTTCTTTGCCTTCTGGGTGATTGTTCACTTGTACCCT 3252

        2838 TTCCTTAAAGGTCTAATGGGTCGCCAAAACCTAACACCAACCATTGTTGTTCTCTGGTCA 2897
             ||||| || ||| | |||||||| || ||||  || || |||||||||  |||||||
        3253 TTCCTCAAGGGTTTGATGGGTCGACAGAACCGGACTCCTACCATTGTTGTGGTCTGGTCT 3312

        2898 GTGCTGTTGGCCTCTGTCTTCTCTCTCGTTTGGGTCAAGATCAATCCATTCGTTAACAAA 2957
             || || ||||| ||| |||||||  |  | ||||| | |||  |||| |||  || | |
        3313 GTTCTCTTGGCTTCTATCTTCTCGTTGTTGTGGGTTAGGATTGATCCCTTCACTAGCCGA 3372

        2958 GTTGATAACACCTTGGTTGCGGA 2980
             ||  | | |     ||  |||
        3373 GTCACTGGCCCGGACATTCTGGA 3395


SULT 4
-09-221-013A-5
 Sequence 5, Application US/09221013A
 Patent No. 6495740
 GENERAL INFORMATION:
  APPLICANT:  Arioli, Antonio
  APPLICANT:  Williamson, Richard E.
  APPLICANT:  Betzner, Andreas S.
  APPLICANT:  Peng, Liangcai
  TITLE OF INVENTION:  Manipulation of cellulose and/or Beta-1,4-glucan
  FILE REFERENCE:  96-98
  CURRENT APPLICATION NUMBER: US/09/221,013A
  CURRENT FILING DATE:  1998-12-23
  PRIOR APPLICATION NUMBER:  PCT/AU97/00402
  PRIOR FILING DATE:  1997-06-24
  PRIOR APPLICATION NUMBER:  AU PO0699
  PRIOR FILING DATE:  1996-06-27
  NUMBER OF SEQ ID NOS:  37
  SOFTWARE:  PatentIn Ver. 2.0
 SEQ ID NO 5
    LENGTH: 3603
    TYPE: DNA
    ORGANISM: Arabidopsis thaliana
    FEATURE:
    NAME/KEY: CDS
    LOCATION: (1)..(3243)
S-09-221-013A-5

 Query Match             34.0%;  Score 1097.8;  DB 4;  Length 3603;
 Best Local Similarity   67.6%;  Pred. No. 8.3e-295;
 Matches 1636;  Conservative    0;  Mismatches  762;  Indels   21;  Gaps    6;

y         551 TGTTTATCCAATTCCACGCAACAAGCTCACACCATACAGAGCAGTGATCATTATGCGACT 610
              |||    || || |||           || || || ||  | |  ||||| ||| | || ||
b         786 TGTGGTGCCTATCCCATCTTCTCGCCTAACCCCTTATCGGGTTGTGATTATTCTCCGGCT 845

y         611 GGTCATTCTGGGCCTCTTCTTCCACTTCAGAATAACAAATCCTGTCGATAGTGCCTTTGG 670
              ||||  || |  |||||||| || |   | | |||  | |||||  | | ||| | |
b         846 TATCATCTTGTGTTTCTTCTTGCAATATCGTACAACTCACCCTGTGAAAAATGCATATCC 905

y         671 CCTGTGGCTTACTTCTGTCATATGTGAGATCTGGTTTGCATTTTCTTGGGTGTTGGATCA 730
              |||||  | || || || || ||||||||||||||||||||||||||||  |  | |||||
b         906 TTTGTGGTTGACCTCGGTTATCTGTGAGATCTGGTTTGCATTTTCTTGGCTTCTTGATCA 965

y         731 GTTCCCCAAGTGGAATCCTGTCAATAGAGAAACGTATATCGAAAGGCTGTCGGCAAGGTA 790
              ||| ||||| ||| | ||  | || || || || ||| | ||  | ||  |  ||| ||
b         966 GTTTCCCAAATGGTACCCCATTAACAGGGAGACTTATCTTGACCGTCTCGCTATAAGATA 1025
```

```
Qy        791 TGAAAGAGAGGGTGA[illegible]TTGTGGATTTTTTCGTGAGTACTGTTGA 850
              |||  |||. || | || || |   || ||| ; || || |||| .| ||
Db       1026 TGATCGAGACGGTGAACCATCACAGTCGTTCGTTGATGTGTTTGTTAGTACAGTGGA 1085

Qy        851 TCCGCTGAAGGAAC[illegible]CCATTGATCACTGCCAATACAGTCCTTTCCATCCTTGCTGTGGA 910
              || |||| || ;  ||  |  | | | | |||| || || || ||| ||||||
Db       1086 CCCATTGAAAGAGC[illegible]CCCTTGTTACAGCAAACACAGTTCTCTCGATTCTTTCTGTGGA 1145

Qy        911 CTATCCCGTCGATAAAGTCTCCTGCTAGGTGTCTGATGATGGTGCAGCTATGCTTTCATT 970
              ||| || || |||||  ||| | || | ||||||| |||||||| || ||  ||||||| ||
Db       1146 CTACCCGGTAGATAAAGTAGCCTGTTATGTTTCAGATGATGGTTCAGCTATGCTTACCTT 1205

Qy        971 TGAATCTCTTGTAGAAACAGCTGAGTTTGCAAAGAAGTGGGTTCCGTTCTGCAAAAAATT 1030
              |||||| |||   |||  ||  |||| | | |||.| || .| ||| | ||||||
Db       1206 TGAATCCCTTTCTGAAACGGCTGAGTTTGCAAAGAAATGGGTACCATTTTGCAAGAAATT 1265

Qy       1031 CTCAATTGAACCAA[illegible]AGCACCGGAGTTTTACTTCTCACAGAAAATTGATTACTTGAAAGA 1090
              |   |||.|||' |  | || |  || .| ||  | |||| || | |||.|||| ||
Db       1266 CAACATTGAACCTA[illegible]GGCCCTGAATTCTATTTTGCCAGAAGATAGATTACTTGAAGGA 1325

Qy       1091 CAAGGTTCAA[illegible]GAAA[illegible]TA[illegible]AATGAAAAGGGATTATGAAGAGTA 1150
              |||| | |||   |  ||  ||  |||  |||  ||  ||  |||  |||
Db       1326 CAAGATCCAA[illegible]TTAAAGA[illegible]TAC[illegible]ATGAAGAGAGAGTATGAAGAGTT 1385

Qy       1151 CAAAGTCCGA[illegible]AAAAAAAAGTTAGAAAACACCTGAAGAAAGGATGGAC 1210
              |.|| |  | |'||  || || ||  ||  ||||  |||.|||||. ||||
Db       1386 TAAAGTGAGGATAAATGCTCTTGTTGCCAAAGCACAGAAAATCCCTGAAGAAGGCTGGAC 1445

Qy       1211 TATGCAAGATGGAACACCTTGGCCTGGGAATAACACACGTGATCACCCTGGGCATGATTC 1270
              ||||; |||;| |  || ||||||' || . ||| | ||||| ||| || ||||, .
Db       1446 AATGCAGGATGGTACTCCCTGGCCTGGTAACAACACTAGAGATCATCCT-GGAATGATAC 1504

Qy       1271 AGGTCTTCCTTGGGAAATACTGGGAGCTCGTGACATTGAAGGAAATGAACTACCTCGTCT 1330
              |.|| | | |  .  :| | |  || |||||| || | ||  | |
Db       1505 AGGTGTTCTTAGGCCA  TAGTGGGGGTCTGGATACCGATGGAAATGAGCTGCCTAGACT 1562

Qy       1331 AGTATATGTCTCCAGGGAGAAGAGACCTGGCTACAAGCACCACAAAAAGGCTGGTGCAGA 1390
              | ||||| || | ;| ||| , |||', |  , ||||,|||||,|||| ||
Db       1563 CATCTATGTTTCTCGTGAAAAGCGGCCTGGATTCAACACCACAAAAAGGCTGGAGCTAT 1622

Qy       1391 AAATGCTCTGGTGAGAGTGTCTGCAGTACTCAAAATGCTCCCTACATCCTCAATGTTGA 1450
              |||||  || |  | |: |||  ||  : ||  ||  | ||  ,  | |; ,| ||
Db       1623 GAATGCATTGATCCGTGTATCTGCTGTTCTTA[illegible]CAATGGAGCATATCTTTTGAACGTGGA 1682

Qy       1451 TTGTGATCACTATGT[illegible]AAACAATAGCAAGGCTGTTCGAGAGGCAATGTGCATCCTGATGGA 1510
              |||||||| ||  ' || || |  ||||  ||  .|| || ||||  || ||,||||
Db       1683 TTGTGATCATTACTTTAATAACAGTAAGGCTATTAAAGAAGCTATGTGTTTCATGATGGA 1742

Qy       1511 CCCACAAGTAGGTCGAGATGTATGCTATGTGCAGTTCCCTCAGAGGTTTGATGGCATAGA 1570
              |||     | ||      |    | ;|||||  ;|||||||||  ;  |||||  || || ||
Db       1743 CCCGGCTATTGGAAAGAAGTGCTGCTATGTCCAGTTCCCTCAACGTTTTGACGGTATTGA 1802

Qy       1571 TAAGAGTGATCGCTACGCCAATCGTAACGTAGTTTTCTTTGATGTTAACATGAAAGGGTT 1630
              |  |  |||||  || |||||  | ;   |||  || || ||| ||||||||| |||||
Db       1803 TTTGCACGATCGATATGCCAACAGGAATATAGTCTTTTTCGATATTAACATGAAGGGGTT 1862

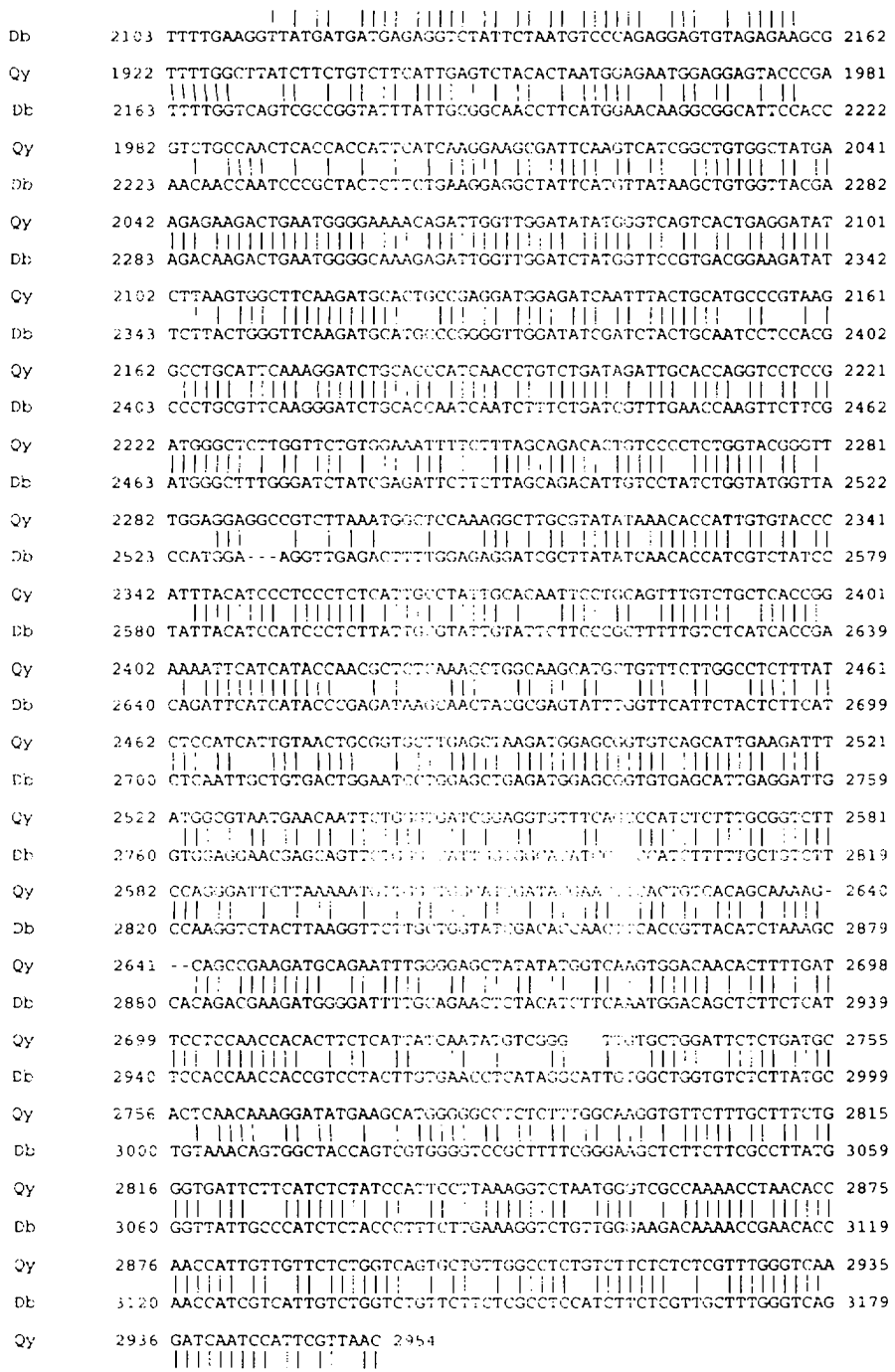
Qy       1631 GGATGGCATTCAAGGACCAGTATACGTAGGAACTGGTTGTGTTTTCAACAGGCAAGCACT 1690
              |||||| || || || |||||||  || ||||||||  ||| || |||||  || || ||
Db       1863 GGATGGTATCCAGGGTCCAGTATATGTGGGTACTGGTTGTTGTTTTAATAGGCAGGCTCT 1922

Qy       1691 TTACGGCTACGGGCGTCCTTCTACGCCCAGCTTACGCAAGAGAAAGGATTCTTCATCCTG 1750
              || || || |  |||  ||  | |                |  |||  |  ;  .
Db       1923 ATATGGGTATGATCCTGTTTTGACGGAAGAAGATTTAGAACCAAATATTATTGTCAAGAG 1982

Qy       1751 CTTCTCATGTTGCTGCCCCCTCAAAGAAGAAGCCTGCTCAAGATCCAGCTGAGGTATACAG 1810
              || |  | |  |    .  || |      ;  ||     .|        ||
Db       1983 CTGTTGCGGGTCAAGGAAGAAAGGTAAAAGTAGCAAGAAGTATAACTACGAAAAGAGGAG 2042

Qy       1811 AGATGCAAAAAAGAGAGGATCTCAATGCTGCCATATTTAATCTTACAGAGATTGATAATTA 1870
              ||      || |||   ||  ||||||, |  ||  ||| |     || || ||| |
Db       2043 AGGCATCAACAGAAGTGACTCCAATGCTCCACTTTTCAATATGGAGGACATCGATGAGGG 2102

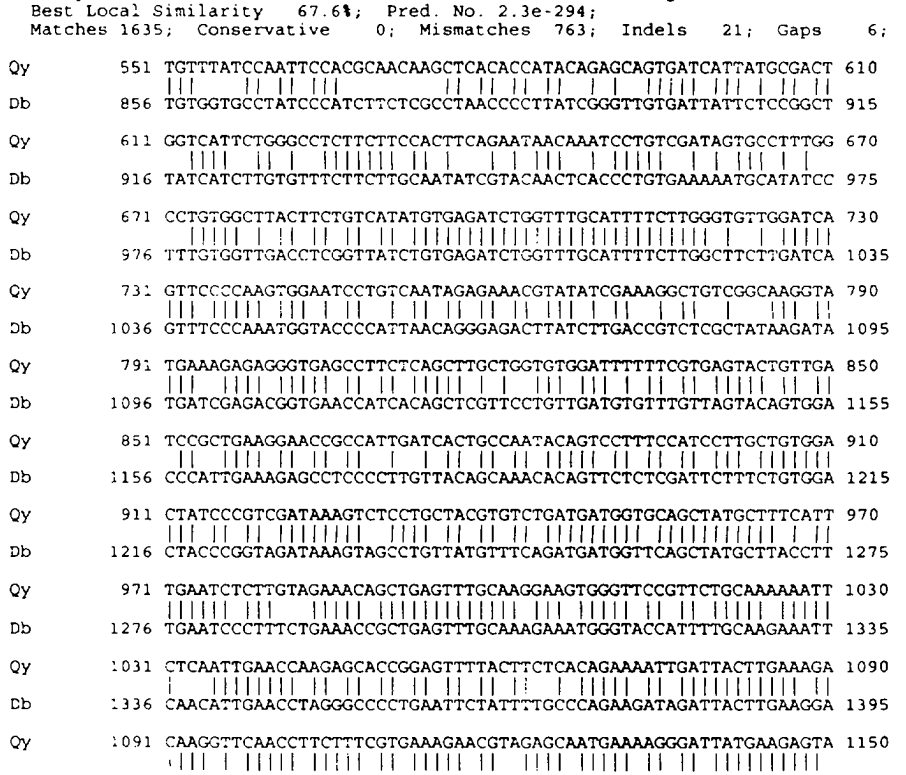
Qy       1871 ---------TGACGAGCATGAAAGGTCAATGCTGATCTCCCAGTTGAGCTTTGAGAAAAC 1921
```

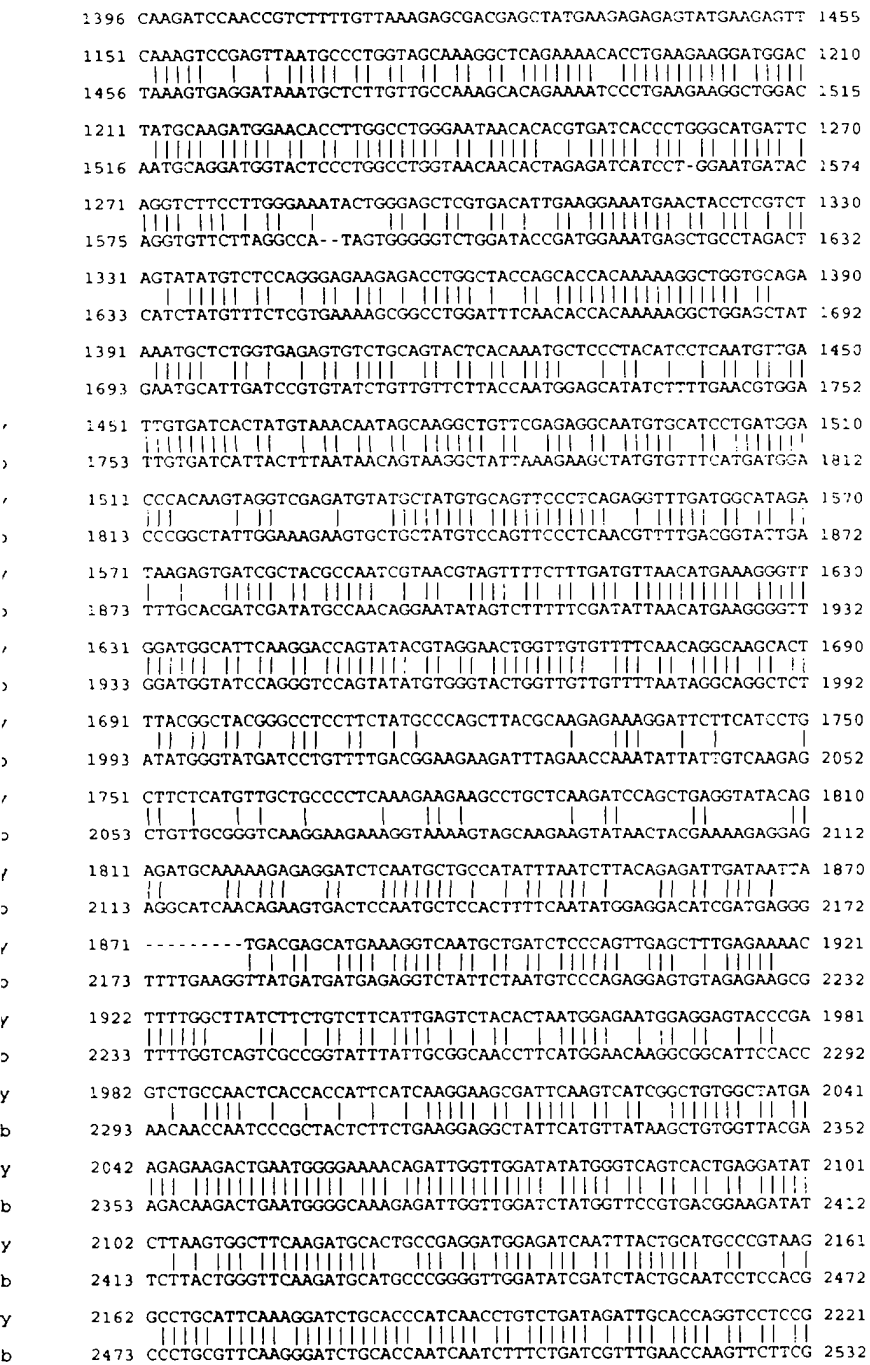


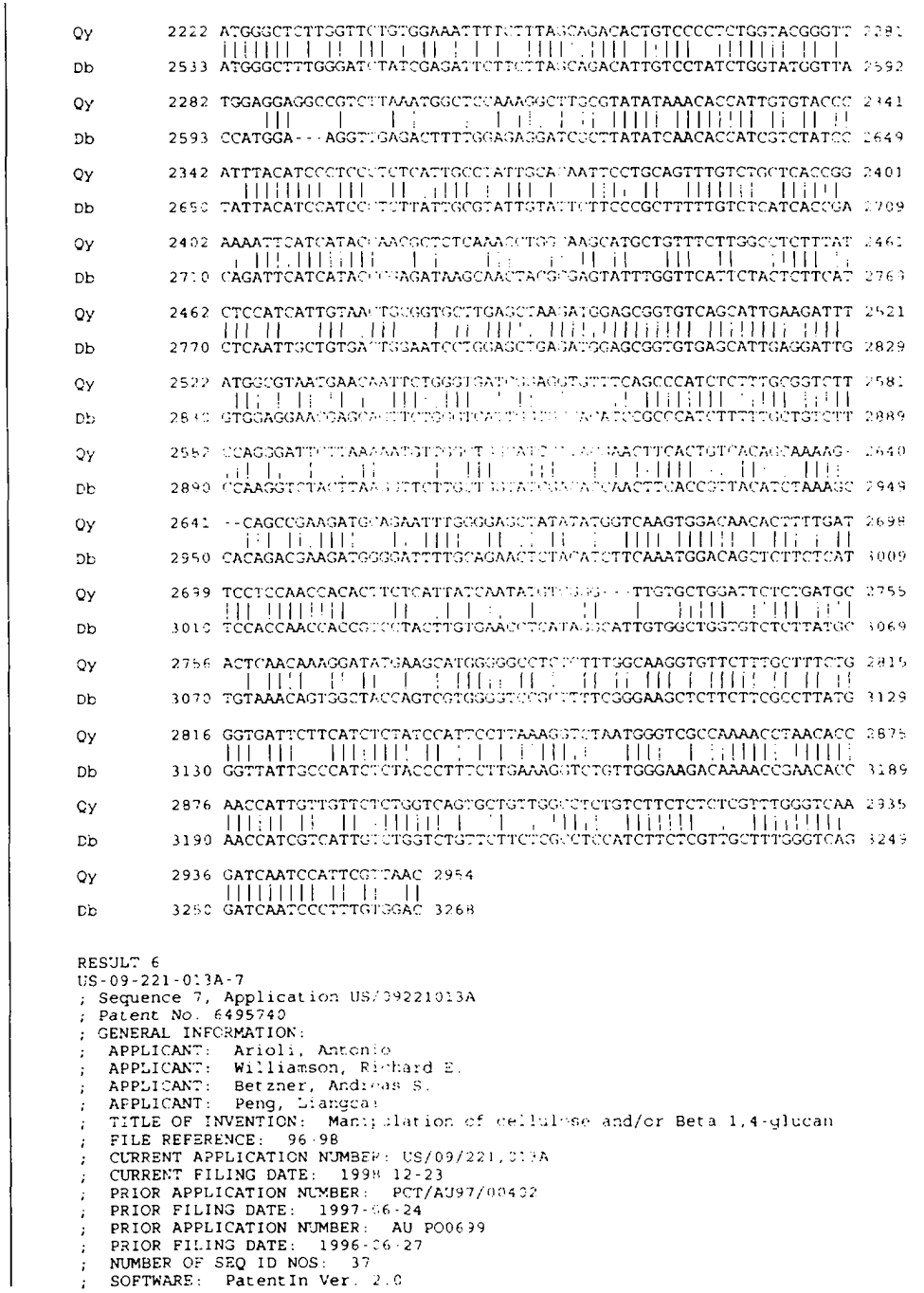
Db 3180 GATCAATCCCTTTGTGGAC 3198

RESULT 5
US-09-221-013A-11
; Sequence 11, Application US/09221013A
; Patent No. 6495740
; GENERAL INFORMATION:
; APPLICANT: Arioli, Antonio
; APPLICANT: Williamson, Richard E.
; APPLICANT: Betzner, Andreas S.
; APPLICANT: Peng, Liangcai
; TITLE OF INVENTION: Manipulation of cellulose and/or Beta-1,4-glucan
; FILE REFERENCE: 96-98
; CURRENT APPLICATION NUMBER: US/09/221,013A
; CURRENT FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: PCT/AU97/00402
; PRIOR FILING DATE: 1997-06-24
; PRIOR APPLICATION NUMBER: AU PO0699
; PRIOR FILING DATE: 1996-06-27
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 3673
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (71)..(3313)
US-09-221-013A-11

Query Match 33.9%; Score 1096.2; DB 4; Length 3673;
Best Local Similarity 67.6%; Pred. No. 2.3e-294;
Matches 1635; Conservative 0; Mismatches 763; Indels 21; Gaps 6;



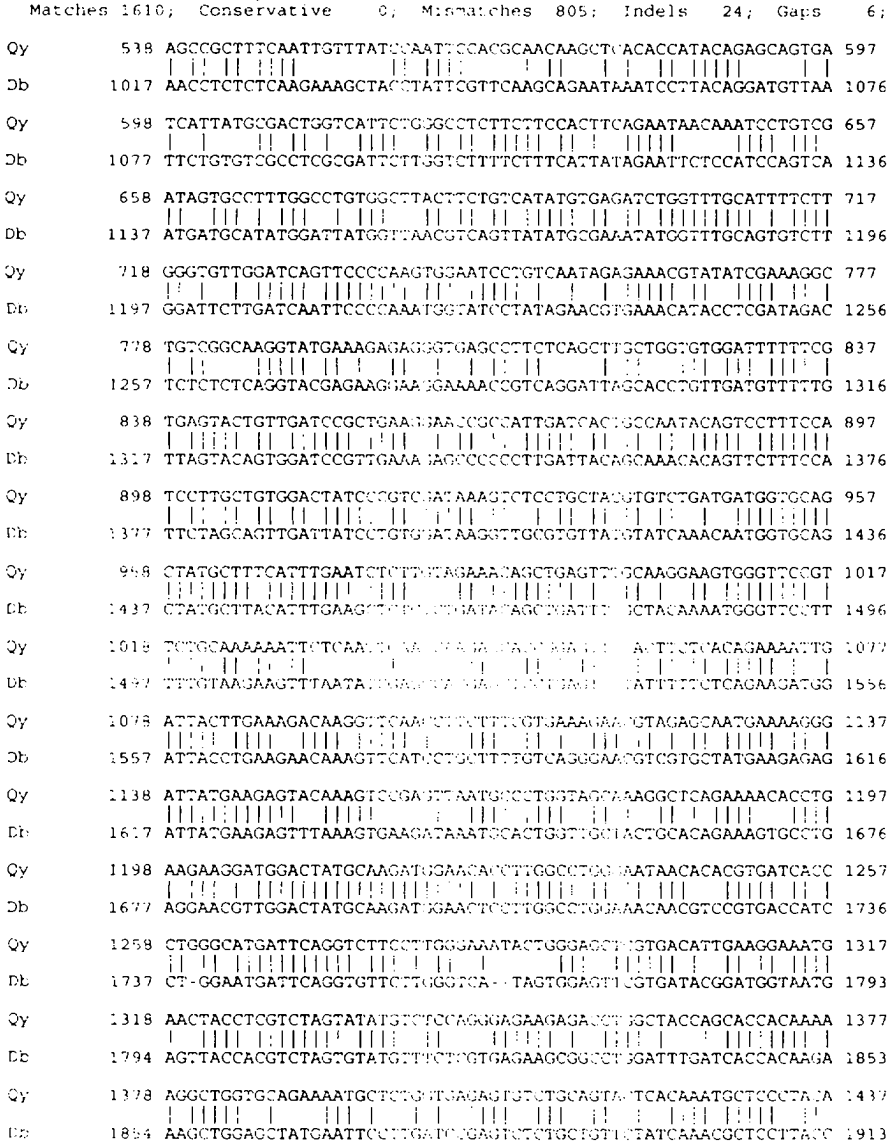


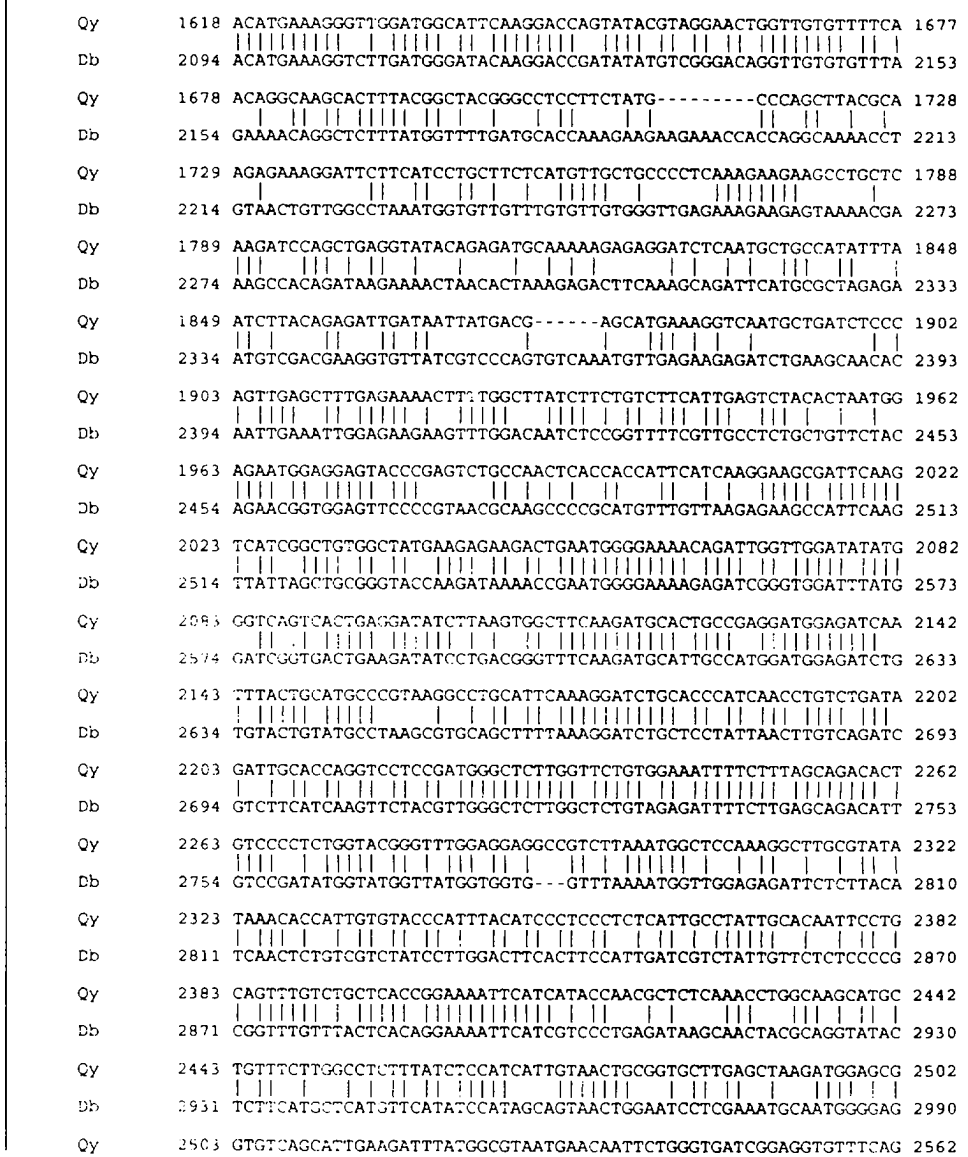


RESULT 6
US-09-221-013A-7
; Sequence 7, Application US/09221013A
; Patent No. 6495740
; GENERAL INFORMATION:
; APPLICANT: Arioli, Antonio
; APPLICANT: Williamson, Richard E.
; APPLICANT: Betzner, Andreas S.
; APPLICANT: Peng, Liangcai
; TITLE OF INVENTION: Manipulation of cellulose and/or Beta 1,4-glucan
; FILE REFERENCE: 96-98
; CURRENT APPLICATION NUMBER: US/09/221,013A
; CURRENT FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: PCT/AU97/00432
; PRIOR FILING DATE: 1997-06-24
; PRIOR APPLICATION NUMBER: AU PO0699
; PRIOR FILING DATE: 1996-06-27
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 7
; LENGTH: 3828
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (239)..(3490)
US-09-221-013A 7

Query Match 32.3%; Score 1043; DB 4; Length 3828;
Best Local Similarity 66.0%; Pred. No. 1.4e-279;
Matches 1610; Conservative 0; Mismatches 805; Indels 24; Gaps 6;







```
b       2991 GTGTCGGAATCGATGATTGGTGGAGAAACGAGCAGTTTTGGGTAATCGGAGGGGCCTCCT 3050

y       2563 CCCATCTCTTTGCGGTCTTCCAGGGATTCTTAAAAATGTTGGCTGGCATCGATACGAACT 2622

b       3051 CGCATCTATTTGCTCTGTTTCAAGGTTTGCTCAAAGTTCTAGCCGGAGTTAACACGAATT 3110

y       2623 TCACTGTCACAGCAAAAGCAGCCGAAGATGCAGAATTTGGGGAGCTATATATGGTCAAGT 2682

b       3111 TCACAGTCACTTCAAAAGCAGCAGACGATGGAGCTTTCTCTGAGCTTTACATCTTCAAGT 3170

y       2683 GGACAACACTTTTGATTCCTCCAACCACACTTC---TCATTATCAATATGTCGGGTTGTG 2739

b       3171 GGACAACTTTGTTGATTCCTCCGACAACACTTCTGATCATTAACATCATTGGAGTTATTG 3230

y       2740 CTGGATTCTCTGATGCACTCAACAAAGGATATGAAGCATGGGGGCCTCTCTTTGGCAAGG 2799

b       3231 TCGGCGTTTCTGATGCCATTAGCAATGGCTATGACTCATGGGGACCTCTCTTTGGGAGAC 3290

y       2800 TGTTCTTTGCTTTCTGGGTGATTCTTCATCTCTATCCATTCCTTAAAGGTCTAATGGGTC 2859

b       3291 TTTTCTTCGCTCTTTGGGTCATTGTTCATTTATACCCATTCCTCAAGGGAATGCTTGGGA 3350

y       2860 GCCAAAACCTAACACCAACCATTGTTGTTCTCTGGTCAGTGCTGTTGGCCTCTGTCTTCT 2919

b       3351 AGCAAGACAAAATGCCTACGATTATTGTGGTCTGGTCTATTCTTCTAGCTTCGATCTTGA 3410

y       2920 CTCTCGTTTGGGTCAAGATCAATCCATTCGTTAACAAAG 2958

b       3411 CACTCTTGTGGGTCAGAATTAACCCGTTTGTGGCTAAAG 3449


ESULT 7
S-09-221-013A-1
 Sequence 1, Application US/09221013A
 Patent No. 6495740
 GENERAL INFORMATION:
  APPLICANT:  Arioli, Antonio
  APPLICANT:  Williamson, Richard E.
  APPLICANT:  Betzner, Andreas S.
  APPLICANT:  Peng, Liangcai
  TITLE OF INVENTION:  Manipulation of cellulose and/or Beta-1,4-glucan
  FILE REFERENCE:  96-98
  CURRENT APPLICATION NUMBER: US/09/221,013A
  CURRENT FILING DATE:  1998-12-23
  PRIOR APPLICATION NUMBER:  PCT/AU97/00402
  PRIOR FILING DATE:  1997-06-24
  PRIOR APPLICATION NUMBER:  AU PO0699
  PRIOR FILING DATE:  1996-06-27
  NUMBER OF SEQ ID NOS:  37
  SOFTWARE:  PatentIn Ver. 2.0
 SEQ ID NO 1
   LENGTH: 2248
   TYPE: DNA
   ORGANISM: Arabidopsis thaliana
   FEATURE:
   NAME/KEY: CDS
   LOCATION: (1)..(1887)
S-09-221-013A-1

 Query Match           24.0%;  Score 775.4;  DB 4;  Length 2248;
 Best Local Similarity  66.1%;  Pred. No. 2.5e-205;
 Matches 1220;  Conservative    0;  Mismatches  606;  Indels   21;  Gaps    6;

y       1123 GAGCAATGAAAAGGGATTATGAAGAGTACAAAGTCCGAGTTAATGCCCTGGTAGCAAAGG 1182

b          2 GAGCTATGAAGAGAGAGTATGAAGAGTTTAAAGTGAGGATAAATGCTCTTGTTGCCAAAG 61

y       1183 CTCAGAAAACACCTGAAGAAGGATGGACTATGCAAGATGGAACACCTTGGCCTGGGAATA 1242

b         62 CACAGAAAATCCCTGGAGAAGGCTGGACAATGCAGGATGGTACTCCCTGGCCTGGTAACA 121

y       1243 ACACACGTGATCACCCTGGGCATGATTCAGGTCTTCCTTGGGAAATACTGGGAGCTCGTG 1302

Db       122 ACACTAGAGATCAT--GGAATGATA--GTGTCTTTAGGCCA--TAGTGGGGGTCTGG 178

Qy      1303 ACATTGAAGGAAAT-AACTACTCTCGT-TA-TATATGTCTCCAGGGAGAAGAGACCTGGCT 1362

Db       179 ATACCGATGGAAAT-AGCTGCTAGA-T-ATCTATGTTTCTCGTGAAAAGCGGCCTGGAT 238

Qy      1363 ACCAGCACCACAAAAAGGCTGGTGCAGAAAATGCTCTGGTGAGAGTGTCTGCAGTACTCA 1422

Db       239 TTCAACACCACAAAAAGGCTGGAGCTATGAATGCATCGATCCGTGTATCTGCTGTTCTTA 298

Qy      1423 CAAATGCTCTGCTACAGCCTCAATGTTGATTGTGATCACTATGTAAACAATAGCAAGGCTG 1482

Db       299 CCAATGGAGCATAT-TTTTGAAGTTGATTGTGATCATTACTTTAATAACAGTAAGGCTA 356

Qy      1483 TTCGAGAGGCAATGTTCATCTTGATGGAGCTCACAAGTAGGTCGAGATGTATGCTATGTGC 1542

Db       357 TTAAAGAAGCTATGT-TTTCATGATGAA-CGGCTATTGGAAAGAAGTGCTGCTATGTCC 416

Qy      1543 AGTTCCCTCAGAGC-T-ATGATATA-AT-AAGAT-GATCGCTACGCCAATCGTAACGTAG 1602

Db       417 AGTTCCCTCAAGGT-T-AGCTATAT-TC-A-AATCGATATGCCAACAGGAATATAG 470

Qy      1603 TTTTCTTTGAT-TTT-A-AT-AAA-T-T-A-TT-ATTCAAGTACCAATATGTTAGGAA 1662

Db       479 TCTTTTTCGATATTAACATGAA-GGGTT-AT-TTATCCACGGTCCAGTATATGTGGGTA 536

Qy      1663 CTGGTTGTGTTTTCAACAGGCAAGCAGTTT-ACGGCTACGGGCCTCCTTCTATGCCGAGCT 1722

Db       537 CTGGTTGTTGTTTTAATAGGCAGGCTCTAT-TGGTATGATCCTGTTTTGACGGAAGAAG 596

Qy      1723 TACGCAAGAGAAAG-ATTCTTCATCTTGCTTCTCATGTTGCTGCCCCTCAAAGAAGAAGC 1782

Db       597 ATTTAGAACCAAATATTATTGTCAAGAGCTGTTGCGGTCAAGGAAGAAAGGTAAAAGTA 656

Qy      1783 CTGCTCAAGATCCA-TTGAGGTATACAGA-AT-GCAAAAAGAGAGGATCTCAATGCTGCCA 1842

Db       657 GCAAGAAGTATAACTACGAAAAGAGGAGAGG-ATCAACAGAAGTGACTCCAATGCTCCAC 716

Qy      1843 TATTTAATCTTACAGAGATTGA---------TAATTATGACGAGCATGAAAGGTCAATGC 1893

Db       717 TTTTCAATATGGAG-ACATCGATGAGCGTTTT-AAGTTATGATGATGAGAGGTCTATTC 774

Qy      1894 TGATCTCCCAGTTGAGCTTTGAGAAAACTTTTGGCTTATCTTCTGTCTTCATTGAGTCTA 1953

Db       775 TAATGTCCCAGAGGAGTGTAGAGAAGCGTTTTGGTCAGTCGCCGGTATTTATTGCGGCAA 834

Qy      1954 CACTAATGGAGAATGGAGGAGTACCCGAGTCTGCCAACTCACCACCATTCATCAAGGAAG 2013

Db       835 CCTTCATGGAACAAGCGGCATTCCACCAACAACCAATCCCGCTACTCTTCTGAAGGAGG 894

Qy      2014 CGATTCAAGTCATCGGCTGTGGCTATGAAGAGAAGACTGAATGGGGAAAACAGATTGGTT 2073

Db       895 CTATTCATGTTATAAGCTGTGGTTACGAAGACAAGACTGAATGGGGCAAAGAGATTGGTT 954

Qy      2074 GGATATATGGGTCAGTCACTGAGGATATCTTAAGTGGCTTCAAGATGCACTGCCGAGGAT 2133

Db       955 GGATCTATGGTTCCGTGACGGAAGATATTTTTACTGGGTTCAAGATGCATGCCCGGGGTT 1014

Qy      2134 GGAGATCAATTTACTGCATGCCCGTAAGGTCTGCATTCAAAGGATCTGCACCCATCAACC 2193

Db       1015 GGATATCGATCTACTGCAATCCTCCACGCCTGCGTTCAAGGGATCTGCACCAATCAATC 1074

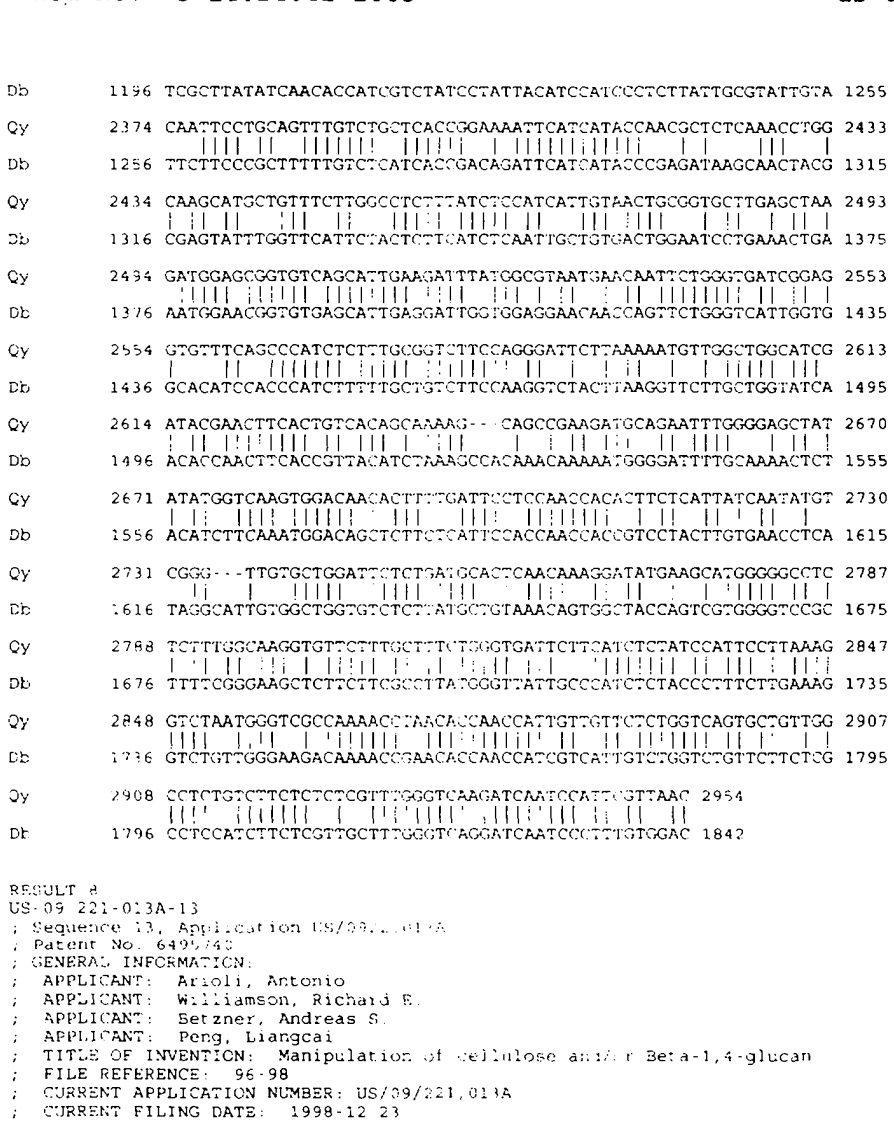
Qy      2194 TGTCTGATAGATTGCACCAGGTCCTCCGATGGGCTCTTGGTTCTGTGGAAATTTTCTTTA 2253

Db       1075 TTTCTGATCGTTTGAACCAAGTTCTTCGATGGGCTTTGGGATCTATCGAGATTCTTCTTA 1134

Qy      2254 GCAGACACTGTCCCCTCTGGTACGGGTTTGGAGGAGCCCGTCTTAAATGGCTCCAAAGGC 2313

Db       1135 GCAGACATTGTCCTATCTGGTATGGTTACCATGGA---AGGTTGAGACTTTTGGAGAGGA 1191

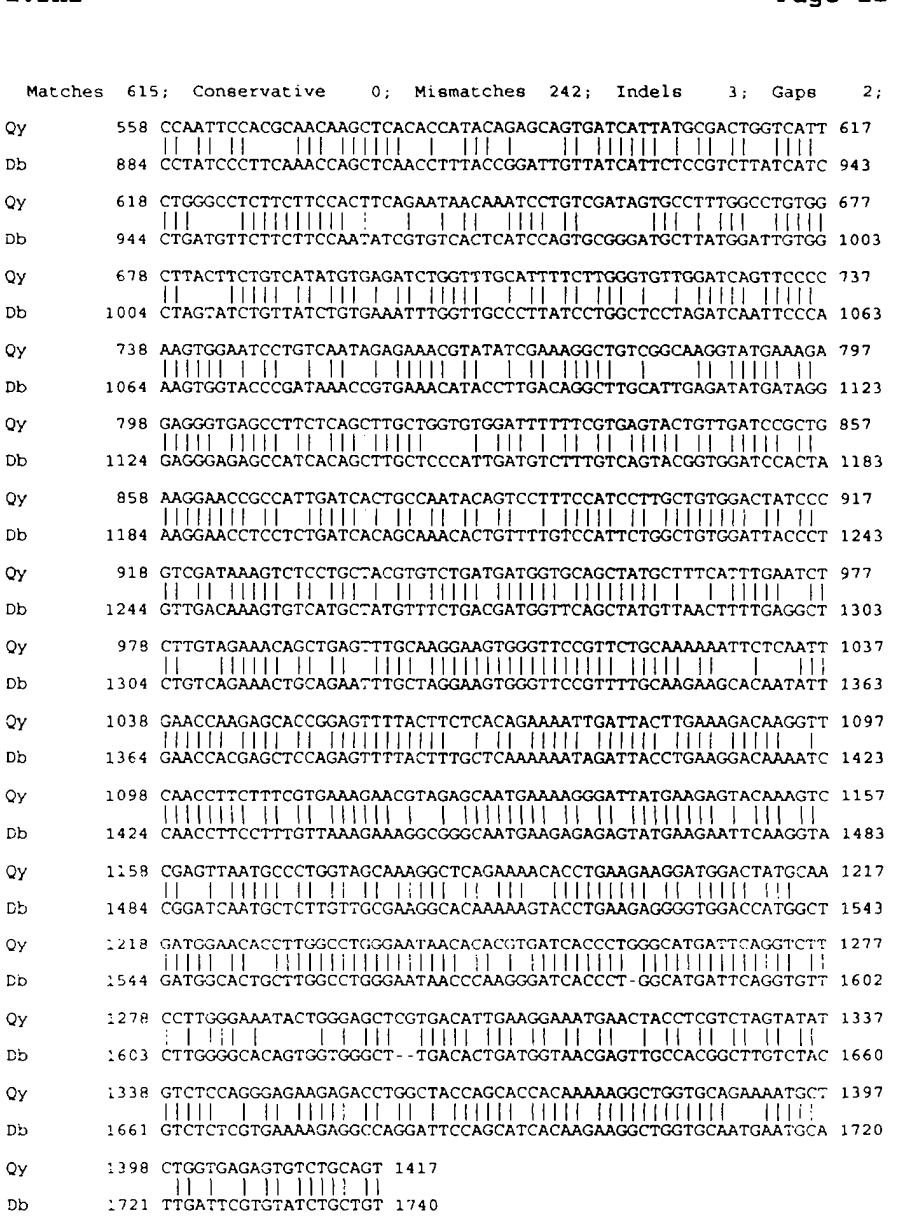
Qy      2314 TTGCGTATATAAACACCATTGTGTACCCATTTACATCCCTCCCTCTCATTGCCTATTGCA 2373
```



RESULT 8
US-09-221-013A-13
; Sequence 13, Application US/09...013A
; Patent No. 6495740
; GENERAL INFORMATION:
; APPLICANT: Arioli, Antonio
; APPLICANT: Williamson, Richard E.
; APPLICANT: Betzner, Andreas S.
; APPLICANT: Peng, Liangcai
; TITLE OF INVENTION: Manipulation of cellulose and/or Beta-1,4-glucan
; FILE REFERENCE: 96-98
; CURRENT APPLICATION NUMBER: US/09/221,013A
; CURRENT FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: PCT/AU97/00400
; PRIOR FILING DATE: 1997-06-24
; PRIOR APPLICATION NUMBER: AU PO3609
; PRIOR FILING DATE: 1996-06-27
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 1741
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (101)..(1741)
US-09-221-013A-13

Query Match 13.6%; Score 446.8; DB 4; Length 1741;
Best Local Similarity 71.5%; Pred. No. 4.3e-114;

Matches 615; Conservative 0; Mismatches 242; Indels 3; Gaps 2;



RESULT 9
US-09-221-013A-3
; Sequence 3, Application US/09221013A
; Patent No. 6495740
; GENERAL INFORMATION:
; APPLICANT: Arioli, Antonio
; APPLICANT: Williamson, Richard E.
; APPLICANT: Betzner, Andreas S.
; APPLICANT: Peng, Liangcai
; TITLE OF INVENTION: Manipulation of cellulose and/or Beta-1,4-glucan

FILE REFERENCE: 96-98
CURRENT APPLICATION NUMBER: US/09/221,013A
CURRENT FILING DATE: 1998-12-23
PRIOR APPLICATION NUMBER: PCT/AU97/00402
PRIOR FILING DATE: 1997-06-24
PRIOR APPLICATION NUMBER: AU PO0699
PRIOR FILING DATE: 1996-06-27
NUMBER OF SEQ ID NOS: 37
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 3
LENGTH: 8411
TYPE: DNA
ORGANISM: Arabidopsis thaliana
-09-221-013A-3

Query Match 10.9%; Score 352.8; DB 4; Length 8411;
Best Local Similarity 64.9%; Pred. No. 1.2e-87;
Matches 635; Conservative 0; Mismatches 252; Indels 91; Gaps 4;

```
Qy       2065 AGATTGGTTGGATATATGGGTCAGTCACTGAGGATATCTTAAGTGGCTTCAAGATGCACT 2124
             |||||||||||| ||||| || || || || |||||  | | ||| ||||||||||||
Db       6653 AGATTGGTTGGATCTATGGTTCCGTGACGGAAGATATTCTTACTGGGTTCAAGATGCATG 6712

Qy       2125 GCCGAGGATGGAGATCAATTTACTGCATGCCCGTAAGGCCTGCATTCAAAGGATCTGCAC 2184
              ||| || |||| ||| || |||||||  ||   ; | ||||| ||||| ||||||||||
Db       6713 CCCGGGGTTGGATATCGATCTACTGCAATCCTCCACGCCCTGCGTTCAAGGGATCTGCAC 6772

Qy       2185 CCATCAACCTGTCTGATAGATTGCACCAGGTCCTCCGATGGGCTCTTGGTTCTGTGGAAA 2244
              | ||||| || |||||| | ||| |||| || || ||||||||| | || ||| | || |
Db       6773 CAATCAATCTTTCTGATCGTTTGAACCAAGTTCTTCGATGGGCTTTGGGATCTATCGAGA 6832

Qy       2245 TTTTCTTTAGCAGACACTGTCCCCTCTGGTACGGGTTTGGAGGAGGCCGTCTTAAATGGC 2304
              || |  |||||||||| |||||  ||||||| || |      |||    |  | | |
Db       6833 TTCTTCTTAGCAGACATTGTCCTATCTGGTATGGTTACCATGGA---AGGTTGAGACTTT 6889

Qy       2305 TCCAAAGGCTTGCGTATATAAACACCATTGTGTACCCATTTACATCCCTCCCTCTCATTG 2364
              |  | ||| | || ||||| |||||||| || || ||  |||||||| ||||||| ||||
Db       6890 TGGAGAGGATCGCTTATATCAACACCATCGTCTATCCTATTACATCCATCCCTCTTATTG 6949

Qy       2365 CCTATTGCACAATTCCTGCAGTTTGTCTGCTCACCGGAAAATTCATCATACCAACGCTCT 2424
              | ||||| |   |||| || ||||||| |||||  | |||||||||||    | | |
Db       6950 CGTATTGTATTCTTCCCGCTTTTTGTCTCATCACCGACAGATTCATCATACCCGAGGTTT 7009

Qy       2425 ------------------------------------------------------------ 2424

Db       7010 GTAAAACTGACCACACTGCTATTTACTATTTGAATCCCATTTTGTGAATGCATTTTTTTG 7069

Qy       2425 ----------------------CAAACCTGGCAAGCATGCTGTTTCTTGGCCTCTTTATC 2462
                                     |||  || || ||  |||  ||  ||||| |||
Db       7070 TCATCATCATTGTTGCAGATAAGCAACTACGCGAGTATTTGGTTCATTCTACTCTTCATC 7129

Qy       2463 TCCATCATTGTAACTGCGGTGCTTGAGCTAAGATGGAGCGGTGTCAGCATTGAAGATTTA 2522
              || ||   ||| ||||   | || ||||| |||||||||||||| |||||||| ||||
Db       7130 TCAATTGCTGTGACTGGAATCCTGGAGCTGAGATGGAGCGGTGTGAGCATTGAGGATTGG 7189

Qy       2523 TGGCGTAATGAACAATTCTGGGTGATCGGAGGTGTTTCAGCCCATCTCTTTGCGGTCTTC 2582
              ||| | || || || |||||||| || || ||    || |||||||| ||||| ||||||
Db       7190 TGGAGGAACGAGCAGTTCTGGGTCATTGGTGGCACATCCGCCCATCTTTTTGCTGTCTTC 7249

Qy       2583 CAGGGATTCTTAAAAATGTTGGCTGGCATCGATACGAACTTCACTGTCACAGCAAAAG-- 2640
              || ||  |  | ||  |  | ||||| ||||| || |||||||| || ||| | ||||
Db       7250 CAAGGTCTACTTAAGGTTCTTGCTGGTATCGACACCAACTTCACCGTTACATCTAAAGCC 7309

Qy       2641 -CAGCCGAAGATGCAGAATTTGGGGAGCTATATATGGTCAAGTGGACAACACTTTTGATT 2699
              ||| ||||||||  || |||| || || || ||  |||| |||||| | ||| | |||
Db       7310 ACAGACGAAGATGGGGATTTTGCAGAACTCTACATCTTCAAATGGACAGCTCTTCTCATT 7369

Qy       2700 CCTCCAACCACACTTCTCATTATCAATATGTCGGG---TTGTGCTGGATTCTCTGATGCA 2756
              || ||||||||  | ||  || | ||  |    ||   |   |||||  ||||| ||||
Db       7370 CCACCAACCACCGTCCTACTTGTGAACCTCATAGGCATTGTGGCTGGTGTCTCTTATGCT 7429

Qy       2757 CTCAACAAAGGATATGAAGCATGGGGGCCTCTCTTTGGCAAGGTGTTCTTTGCTTTCTGG 2816
              | ||||  || ||  |  | ||||| || || || || ||| | ||||| || || |||
```

```
Db       7430 GTAAACAGTGGCTACTAGTCGTGGGGTCGGCTTTTCGGGAAGCTCTTCTTCGCCTTATGG 7489

Qy       2817 GTGATTCTTCATCTCTATCCATTCCTTAAAGGTCTAATGGGTCGCCAAAACCTAACACCA 2876
              || |||   ||||||  || ||| | | |||||  ||||  | ||.|||  ||||.|
Db       7490 GTTATTGCCCATCTCTACCCTTTCTTGAAAGGTCTGTTGGGAAGACAAAACCGAACACCA 7549

Qy       2877 ACCATTGTTGTTCTTTGGTCAGTGCTGTTGGCCTCTGTCTTCTCTCTCGTTTGGGTCAAG 2936
              ||||| ||  || | |||| || | | |||||  |||||||  |  |||||||| |
Db       7550 ACCATCGTCATTGTTTGGTCTGTTCTTCTGGCCTCCATCTTCTCGTTGCTTTGGGTCAGG 7609

Qy       2937 ATCAATCCATTCGTTAAC 2954
              |||||||  || ||  ||
Db       7610 ATCAATCCCTTTGTTGAC 7627
```

RESULT 10
US-09-221-013A-4
; Sequence 4, Application US/09/221,013A
; Patent No. 6495740
; GENERAL INFORMATION:
; APPLICANT: Arioli, Antony
; APPLICANT: Williamson, Richard E
; APPLICANT: Betzner, Andreas S
; APPLICANT: Peng, Lianguo
; TITLE OF INVENTION: Manipulation of Cellulose and/or Beta 1,4-Glucan
; FILE REFERENCE: 96-98
; CURRENT APPLICATION NUMBER: US/09/221,013A
; CURRENT FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: PCT/AU97/00402
; PRIOR FILING DATE: 1997-06-24
; PRIOR APPLICATION NUMBER: AU PO0699
; PRIOR FILING DATE: 1996-06-27
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 5009
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(5009)
; OTHER INFORMATION: N is A, T, G or C.
US-09-221-013A-4

Query Match 6.0%; Score 194.6; DB 4; Length 5009;
Best Local Similarity 72.4%; Pred. No. 8.5e-44;
Matches 252; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

```
Qy        785 AAGGTATGAAAGAGAGGGTGAGCCTTCTCAGCTTGCTGGTGTGGATTTTTTCGTGAGTAC 844
              |||||| || |  || || |  || ||      | ||  ||| |||| || ||||||
Db       2773 AAGGTACGAGAAGGAAGGAAAACCGTCAGGATTAGCACCTGTTGATGTTTTTGTTAGTAC 2832

Qy        845 TGTTGATCCGCTGAAGGAACCGCCATTGATCACTGCCAATACAGTCCTTTCCATCCTTGC 904
              || |||||| ||| || || || ||||  | ||  | |||| |||||||| || ||
Db       2833 AGTGGATCCGTTGAAAGAGCCACCCTTGATTACAGCAAACACAGTTCTTTCCATTCTAGC 2892

Qy        905 TGTGGACTATCCCGTGGATAAAGTCTCCTGCTACGTGTCTGATGATGGTGCAGCTATGCT 964
              .| || |||||  ||  |||  || ' ||  ' ||  .  | ||||||||||||||||
Db       2893 AGTTGATTATCCTGTGGATAAGGTTGCGTGTTATGTATCAAACAATGGTGCAGCTATGCT 2952

Qy        965 TTCATTTGAATCTCTGTAGAAACAGCTGAGTTTGCAAGGAAGTGGGTTCCGTTCTGCAA 1024
              | |||||||  |||     |  ||| |||  |||| || '' |||||||| || || ||
Db       2953 TACATTTGAAGCTCTTCTGATACAGCTGAGTTTGCTAGAAAATGGGTTCCTTTTTGTAA 3012

Qy       1025 AAAATTCTCAATTGAACCAAGAGCACGGGAGTTTTACTTCTCACAGAAAATTGATTACTT 1084
              || ||    ||  || ||| |||| .. |||'  || || || |||||  | ||| || |
Db       3013 GAAGTTTAATATCGAGCCACGAGCTCCTGAGTGGTATTTTTCTCAGAAGATGGATTACCT 3072

Qy       1085 GAAAGACAAGGTTCAACCTTCTTTCGTGAAAGAACGTAGAGCAATGAA 1132
              |||  |||  ||| | ||' |||  ||  ;  ||| || | || ||||
Db       3073 GAAGAACAAAGTTCATCCTGCTTTTGTCAGGGAACGTCGTGCTATGAA 3120
```

```
RESULT 11
US 09-313-294A-5519
; Sequence 5519, Application US/09313294A
; Patent No. 6476212
; GENERAL INFORMATION:
;  APPLICANT:  Lalgudi, Raghunath V.
;  APPLICANT:  Ito, Laura Y.
;  APPLICANT:  Sherman, Bradley K.
;  TITLE OF INVENTION:  POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
;  FILE REFERENCE:  PL-0017 US
;  CURRENT APPLICATION NUMBER: US/09/313,294A
;  CURRENT FILING DATE:  1999-05-14
;  NUMBER OF SEQ ID NOS:  7600
;  SOFTWARE:  PERL Program
; SEQ ID NO 5519
;   LENGTH: 291
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. 6476212 7003502858H1
;   NAME/KEY: unsure
;   LOCATION: 6, 40, 54, 65, 67, 200, 221, 230, 238, 241
;   OTHER INFORMATION: a, t, c, g, or other
US 09-313-294A-5519

  Query Match             4.3%;  Score 138.2;  DB 4;  Length 291;
  Best Local Similarity   70.5%;  Pred. No. 9.9e-29;
  Matches  206;  Conservative    0;  Mismatches   82;  Indels    4;  Gaps    2;

Qy       2127 CGAGGATGGAGATCAATTTACTGCATGCCCGTAAGGCCTGCATTCAAAGGATCTGCACCC 2186
             ||| | ||| ||||||| |||| |||||     | |   |  |||| || ||| ||
Db          3 CGANGCTGGCGATCAATCTACTGCATGCCTAAGCGACNAGCTTTCAAGGGANCTGCTCCT 62

Qy       2187 AT-CAACCTGTCTGATAGATTGCACCAGGTGCTCCGATGGGCTCTTGGTTCTGTGGAAAT 2245
              || | |||| || ||| | ||| | || | || | || ||||||||||  | |||||
Db         63 ATNCNACCTTTCGGATCGTTTGAATCAAGTGCTTCGGTGGGCTCTTGGTTCCATTGAAAT 122

Qy       2246 TTTCTTTAGCAGACACTGTCCCCTCTGGTACGGGTTTGGAGGAGGCCGTCTTAAATGGCT 2305
              | || ||'|| || |||||   ||||| || | |  ||| || ||||||  |
Db        123 TCTTTTCAGCAGGCATTGTCCCATATGGTATGGCTAT---GGAGGTCGGCTTAAATTCCT 179

Qy       2306 CCAAAGGCTTGCGTATATAAAAA AGTGT TACCCATTTA ATCCTTCCCTCTCATTGC 2365
              | | | ||| | | |  |     ||   | |  ||  ||  ||  |
Db        180 GGATAGATTTGCTTATATCAACA AA AAT TATCCACT ANATCAATCCNGCTCCTCNT 239

Qy       2366 CTATTGCACAATTCCTGCAGTTTGTTGCTTACCGGAAAATCCATCATACCA 2417
              || |||| | | || |||| || || || | |  ||  ||| || |..
Db        240 GTACTGCATATTGCCAGCAGTNTGTCTTCTAACTGGGAAGTCCATCATCCCA 291


RESULT 12
US-09-313-294A-6485
; Sequence 6485, Application US/09313294A
; Patent No. 6476212
; GENERAL INFORMATION:
;  APPLICANT:  Lalgudi, Raghunath V.
;  APPLICANT:  Ito, Laura Y.
;  APPLICANT:  Sherman, Bradley K.
;  TITLE OF INVENTION:  POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
;  FILE REFERENCE:  PL-0017 US
;  CURRENT APPLICATION NUMBER: US/09/313,294A
;  CURRENT FILING DATE:  1999-05-14
;  NUMBER OF SEQ ID NOS:  7600
;  SOFTWARE:  PERL Program
; SEQ ID NO 6485
;   LENGTH: 281
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. 6476212 7003519828H1
US 09-313-294A-6485

  Query Match             4.2%;  Score 136;  DB 4;  Length 281;
  Best Local Similarity   67.9%;  Pred. No. 4e-28;
  Matches  190;  Conservative    0;  Mismatches   90;  Indels    0;  Gaps    0;

Qy       2322 ATAAACACCATTGTGTACCCATTTACATCCCTCCCTCTCATTGCCTATTGCACAATTCCT 2381
              || |||||||   | |||||  | || |||||||| ||| |    || || |   | |||
Db          2 ATCAACACCACCATCTACCCGCTCACGTCCCTCCCGCTCCTCATTTACTGTATCCTGCCT 61

Qy       2382 GCAGTTTGTCTGCTCACCGGAAAATTCATCATACCAACGCTCTCAAACCTGGCAAGCATG 2441
              ||  | || |||||||| || || |||||||| |||  | ||   ||| | || || ||
Db         62 GCCATCTGCCTGCTCACGGGGAAGTTCATCATCCCAGAGATCAGCAACTTCGCTAGTATC 121

Qy       2442 CTGTTTCTTGGCCTCTTTATCTCCATCATTGTAACTGCGGTGCTTGAGCTAAGATGGAGC 2501
              |||  |    ||||| ||||| ||| | |  || |   | || ||| | || ||||||
Db         122 TGGTTCATCTCTCTCTTCATCTCGATCTTCGCCACGGGTATCCTGGAGATGAGGTGGAGC 181

Qy       2502 GGTGTCAGCATTGAAGATTTATGGCGTAATGAACAATTCTGGGTGATCGGAGGTGTTTCA 2561
              || ||  |||| || || |  ||| | || || || |||||||| ||||||||  | ||
Db         182 GGCGTGGGCATCGACGAGTGGTGGAGGAACGAGCAGTTCTGGGTCATCGGAGGCATCTCC 241

Qy       2562 GCCCATCTCTTTGCGGTCTTCCAGGGATTCTTAAAAATGT 2601
              ||||| ||||| || |||||||||||  || | ||  |||
Db         242 GCCCACCTCTTCGCCGTCTTCCAGGGCCTCCTCAAGGTGT 281


RESULT 13
US-09-313-294A-849
; Sequence 849, Application US/09313294A
; Patent No. 6476212
; GENERAL INFORMATION:
;  APPLICANT:  Lalgudi, Raghunath V.
;  APPLICANT:  Ito, Laura Y.
;  APPLICANT:  Sherman, Bradley K.
;  TITLE OF INVENTION:  POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
;  FILE REFERENCE:  PL-0017 US
;  CURRENT APPLICATION NUMBER: US/09/313,294A
;  CURRENT FILING DATE:  1999-05-14
;  NUMBER OF SEQ ID NOS:  7600
;  SOFTWARE:  PERL Program
; SEQ ID NO 849
;   LENGTH: 268
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. 6476212 7005499401H1
US-09-313-294A-849

  Query Match             4.2%;  Score 135;  DB 4;  Length 268;
  Best Local Similarity   69.6%;  Pred. No. 7.4e-28;
  Matches  183;  Conservative    0;  Mismatches   80;  Indels    0;  Gaps    0;

Qy       2454 CTCTTTATCTCCATCATTGTAACTGCGGTGCTTGAGCTAAGATGGAGCGGTGTCAGCATT 2513
              || |||||| |||  |||  ||     | || ||  | |||||||| |||||  | |||
Db          5 CTTTTTATCTGCATTTTTGCTACGAGCATCCTAGAAATGAGATGGAGTGGTGTTGGAATT 64

Qy       2514 GAAGATTTATGGCGTAATGAACAATTCTGGGTGATCGGAGGTGTTTCAGCCCATCTCTTT 2573
              || || |  ||| | ||||| || |||||||| || |||||||| ||  | || ||||||
Db         65 GATGACTGGTGGAGGAATGAGCAGTTCTGGGTCATTGGAGGTGTGTCCTCACACCTCTTT 124

Qy       2574 GCGGTCTTCCAGGGATTCTTAAAAATGTTGGCTGGCATCGATACGAACTTCACTGTCACA 2633
              || || |||||||| |  | ||  |  | |||||  | |||||  | ||||||  || |||
Db         125 GCTGTGTTCCAGGGACTTCTCAAGGTCATAGCTGGTGTTGATACAAGCTTCACCGTGACA 184

Qy       2634 GCAAAAGCAGCCGAAGATGCAGAATTTGGGGAGCTATATATGGTCAAGTGGACAACACTT 2693
              ||||  |  | || ||||  || ||    ||||||||||   |||| ||||| ||  |
Db         185 TCAAAGGGTGGAGATGATGAGGAGTTCTCAGAGCTATATACATTCAAATGGACTACCTTA 244

Qy       2694 TTGATTCCTCCAACCACACTTCT 2716
              :'||| ||||| |||||  | ||
Db         245 TTGATACCTCCTACCACCTTGCT 267
```

SULT 14
-09-313-294A-4753
Sequence 4753, Application US/09313294A
Patent No. 6476212
GENERAL INFORMATION:
APPLICANT: Lalgudi, Raghunath V.
APPLICANT: Ito, Laura Y.
APPLICANT: Sherman, Bradley K.
TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
FILE REFERENCE: PL-0017 US
CURRENT APPLICATION NUMBER: US/09/313,294A
CURRENT FILING DATE: 1999-05-14
NUMBER OF SEQ ID NOS: 7600
SOFTWARE: PERL Program
SEQ ID NO 4753
LENGTH: 300
TYPE: DNA
ORGANISM: Zea mays
FEATURE:
NAME/KEY: misc_feature
OTHER INFORMATION: Incyte ID No. 6476212 700348982H1
NAME/KEY: unsure
LOCATION: 64, 189
OTHER INFORMATION: a, t, c, g, or other
3-09-313-294A-4753

Query Match 4.0%; Score 129.8; DB 4; Length 300;
Best Local Similarity 67.2%; Pred. No. 2.2e-26;
Matches 197; Conservative 0; Mismatches 94; Indels 2; Gaps 1;

```
/       2420 GCTCTCAAACCTGGCAAGCATGCTGTTTCTTGGCCTCTTTATCTCCATCATTGTAACTGC 2479
            | ||   ||| | || |||||   |||  |   |||||| |||||  ||| | |  || |
>          4 GATCAGCAACTTCGCCAGCATCTGGTTCATCTCCCTCTTCATCTCGATCTTCGCCACGGG 63

/       2480 GGTGCTTGAGCTAAGATGGAGCGGTGTCAGCATTGAAGATTTATGGCGTAATGAACAATT 2539
            | || ||| | || |||||||| ||  |||| || || |  ||| | || || || ||
>         64 NATCCTGGAGATGAGGTGGAGCGGGGTGGGCATCGACGAGTGGTGGAGGAACGAGCAGTT 123

/       2540 CTGGGTGATCGGAGGTGTTTCAGCCCATCTCTTTGCGGTCTTCCAGGGATTCTTAAAAAT 2599
            |||||||||||| ||  | || || || ||||| || || |||||||  |  | ||  |
>         124 CTGGGTGATCGGGGGCATCTCCGCGCACCTCTTCGCCGTGTTCCAGGGCCTGCTCAAGGT 183

y       2600 GTTGGCTGGCATCGATACGAACTTCACTGTCACAGCAAAAGCAGCCGAAGATGCAG--AA 2657
            | |||  |||||||| || |||||||| |||||  | || ||  | || || |  |  |
>         184 GCTGGNCGGCATCGACACCAACTTCACCGTCACCTCCAAGGCCTCGGACGAGGACGGCGA 243

y       2658 TTTGGGGAGCTATATATGGTCAAGTGGACAACACTTTTGATTCCTCCAACCAC 2710
            || | |  || || ||| |||||||||| || ||  | || || || ||||
>         244 CTTCGCGGACTGTACATGTTCAAGTGGACGACGCTCCTTATCCCGCCCACCAC 296
```

ESULT 15
S-09-313-294A-1981
Sequence 1981, Application US/09313294A
Patent No. 6476212
GENERAL INFORMATION:
APPLICANT: Lalgudi, Raghunath V.
APPLICANT: Ito, Laura Y.
APPLICANT: Sherman, Bradley K.
TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
FILE REFERENCE: PL-0017 US
CURRENT APPLICATION NUMBER: US/09/313,294A
CURRENT FILING DATE: 1999-05-14
NUMBER OF SEQ ID NOS: 7600
SOFTWARE: PERL Program
SEQ ID NO 1981
LENGTH: 283
TYPE: DNA
ORGANISM: Zea mays
FEATURE:
NAME/KEY: misc_feature
OTHER INFORMATION: Incyte ID No. 6476212 700551845H1
; NAME/KEY: unsure
; LOCATION: 55, 189
; OTHER INFORMATION: a, t, [illegible]
US 09-313-294A 1981

Query Match [illegible]; Score 127; DB 4; Length 283;
Best Local Similarity [illegible]; Pred. No. [illegible];
Matches 200; Conservative 0; Mismatches 67; Indels 4; Gaps 3;

```
Qy       2210 CCAGGTCCTCCGAT-[illegible]AAATTTTCTTTAGCAGACACTGTCCCCT 2269
             || || ||| |  ||| |  |  || |||| ||  | ||| | || ||
Db          1 CCAAGTTCTCGGGT-[illegible]AGATCTTCTTCAGTCGGCACTGNCCGCT 60

Qy       2270 CTGGTA-GGGTTTG-A-AGGGCGGTCTTAAAT-G-TCCAAAGGCTTGCGTATATAAACAC 2329
             | ||| | |  |'   ||| |  | - ||  |||| |'||  | || || |||||
Db         61 GTGGTACGGTTACG-[illegible]AGAGGCTCTCCTACATCAACAC 120

Qy       2330 CATTGTGTA-CCAC-[illegible]TATTGCACAATTGCTGCAGTTTG 2389
             || |||'.|||  |  |  ||  ||  || |  |  | . ||||  |||
Db        121 CATGGTGTA-CCG-[illegible]TACTGTTGACTGCCTGC-CATTG 179

Qy       2390 TCTGCTGAC[illegible]AAGCTTGCAA-CAC-CTGTTTCT 2440
             || ||| ||| |  |  |  |  |  |  |  ||| ||
Db        180 CCTGCTGAC-[illegible]AACGCTGCAA-GATATGGTTTCT 237

Qy       2440 TGGCCTCTTTATCT-ATCATTGTAACTG-G 2471
             ||||||||| |  |:  ||| ,| |  |||
Db        238 TGGCCTCTTCA-GT-CATCATCGTGAGAG-G 269
```

Search completed: October 30, 2003, 16:43:49
Job time : 176.411 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 30, 2003, 12:47:22 ; Search time 811.873 Seconds
(without alignments)
10827.272 Million cell updates/sec

Title: US-09-980-043-1
Perfect score: 3232
Sequence: 1 gtcgacccacgcgtccgctc.....aagaaaaaaaaaaaaaaaaa 3232

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1811591 seqs, 1359936290 residues

Total number of hits satisfying chosen parameters: 3623182

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1989.8	61.6	3328	10	US-09-838-539-1	Sequence 1, Appli
2	1482.6	45.9	3231	12	US-10-209-059-27	Sequence 27, Appl
3	1152.6	35.7	3517	10	US-09-900-237-13	Sequence 13, Appl
4	1149.8	35.6	3753	12	US-10-160-719-33	Sequence 33, Appl
5	1149.8	35.6	3753	12	US-10-160-719-53	Sequence 53, Appl
6	1149.8	35.6	3780	12	US-10-209-059-1	Sequence 1, Appli
7	1149.8	35.6	3780	12	US-10-160-719-9	Sequence 9, Appli
8	1138.8	35.2	3704	12	US-10-209-059-13	Sequence 13, Appl
9	1138.8	35.2	3704	12	US-10-160-719-57	Sequence 57, Appl
10	1137.6	35.2	3725	12	US-10-209-059-45	Sequence 45, Appl
11	1137.6	35.2	3725	12	US-10-160-719-13	Sequence 13, Appl
12	1137.6	35.2	3725	12	US-10-160-719-21	Sequence 21, Appl
13	1137.6	35.2	3725	12	US-10-160-719-41	Sequence 41, Appl
14	1133.8	35.1	3614	14	US-10-229-193-9	Sequence 9, Appli
15	1133.6	35.1	3626	10	US-09-900-237-29	Sequence 29, Appl
16	1128.6	34.9	3799	12	US-10-267-459-5	Sequence 5, Appli
17	1128.6	34.9	3799	12	US-10-209-059-21	Sequence 21, Appl
18	1106.2	34.2	3936	10	US-09-900-237-9	Sequence 9, Appli
19	1106.2	34.2	3969	12	US-10-209-059-49	Sequence 49, Appl
20	1106.2	34.2	3969	12	US-10-160-719-17	Sequence 17, Appl
21	1106.2	34.2	3969	12	US-10-160-719-37	Sequence 37, Appl
22	1103.2	34.1	2830	12	US-10-267-459-1	Sequence 1, Appli
23	1103.2	34.1	2830	12	US-10-209-059-5	Sequence 5, Appli
24	1101.6	34.1	3746	12	US-10-160-719-29	Sequence 29, Appl
25	1101.6	34.1	3746	12	US-10-160-719-49	Sequence 49, Appl
26	1101.6	34.1	3773	12	US-10-209-059-9	Sequence 9, Appli
27	1101.6	34.1	3773	12	US-10-160-719-5	Sequence 5, Appli
28	1097.8	34.0	3603	14	US-10-229-193-5	Sequence 5, Appli
29	1097	33.9	3568	12	US-10-209-059-41	Sequence 41, Appl
30	1097	33.9	3568	12	US-10-160-719-1	Sequence 1, Appli
31	1097	33.9	3776	10	US-09-900-237-3	Sequence 3, Appli
32	1096.2	33.9	3673	14	US-10-229-193-11	Sequence 11, Appl
33	1082	33.5	3786	10	US-09-900-237-7	Sequence 7, Appli
34	1082	33.5	3813	12	US-10-209-059-17	Sequence 17, Appl
35	1082	33.5	3813	12	US-10-160-719-25	Sequence 25, Appl
36	1082	33.5	3813	12	US-10-160-719-45	Sequence 45, Appl
37	1073	33.2	2890	10	US-09-900-237-17	Sequence 17, Appl
38	1051	32.5	3255	10	US-09-938-842A-559	Sequence 559, App
39	1048.6	32.4	3255	10	US-09-938-842A-2530	Sequence 2530, Ap
40	1043	32.3	3828	14	US-10-229-193-7	Sequence 7, Appli
41	1031.8	31.9	3563	10	US-09-900-237-25	Sequence 25, Appl
42	998.8	30.9	3028	12	US-10-209-059-29	Sequence 29, Appl
43	982	30.4	2663	10	US-09-900-237-23	Sequence 23, Appl
44	935.4	28.9	3470	12	US-10-209-059-25	Sequence 25, Appl
45	775.4	24.0	2248	14	US-10-229-193-1	Sequence 1, Appli

ALIGNMENTS

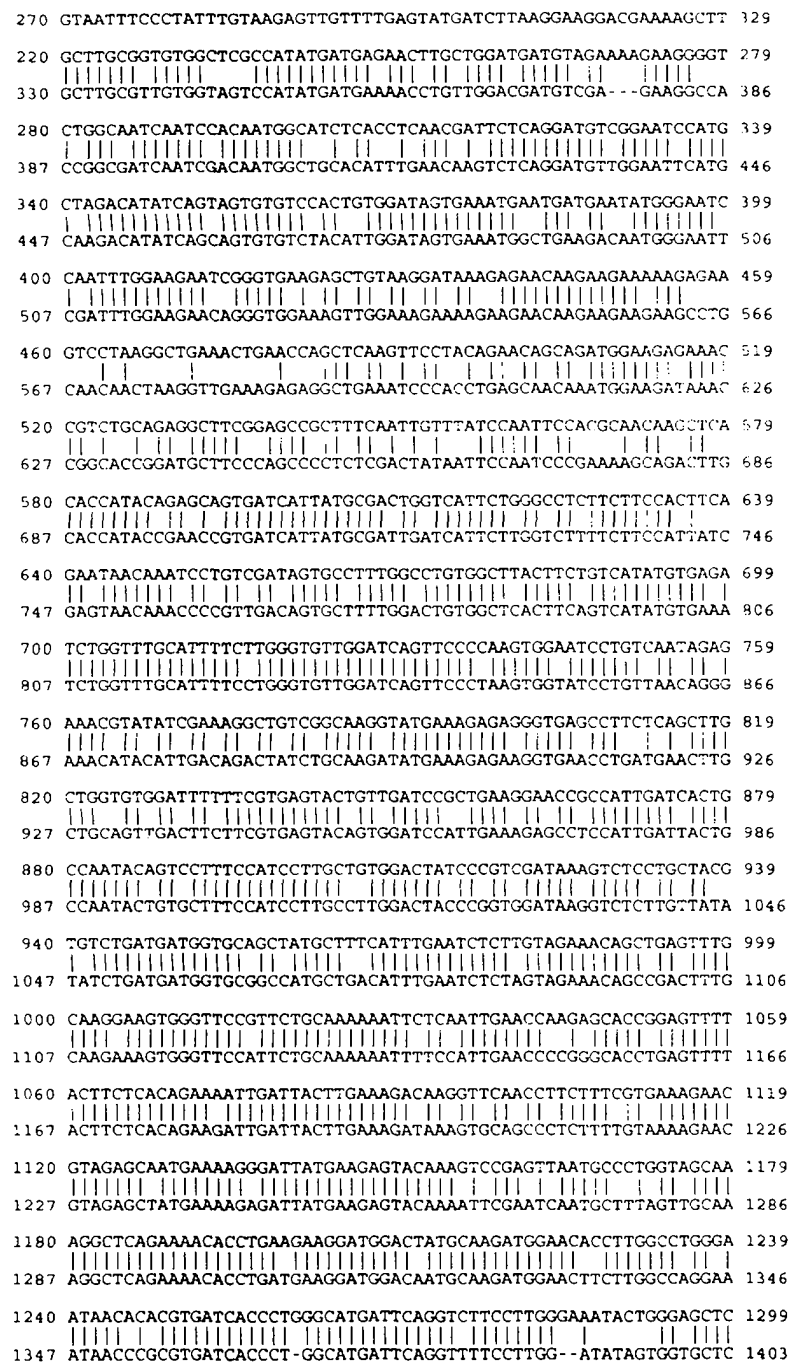
RESULT 1
US-09-838-539-1
; Sequence 1, Application US/09838539
; Patent No. US20020129401A1
; GENERAL INFORMATION:
; APPLICANT: Stalker, D. et al.
; TITLE OF INVENTION: Plant Cellulose Synthase and Promoter
; TITLE OF INVENTION: Sequences
; FILE REFERENCE: 15621/03/US
; CURRENT APPLICATION NUMBER: US/09/838,539
; CURRENT FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/029,987
; PRIOR FILING DATE: 1996-10-29
; PRIOR APPLICATION NUMBER: 08/960,048
; PRIOR FILING DATE: 1997-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 3328
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Oligonucleotide
US-09-838-539-1

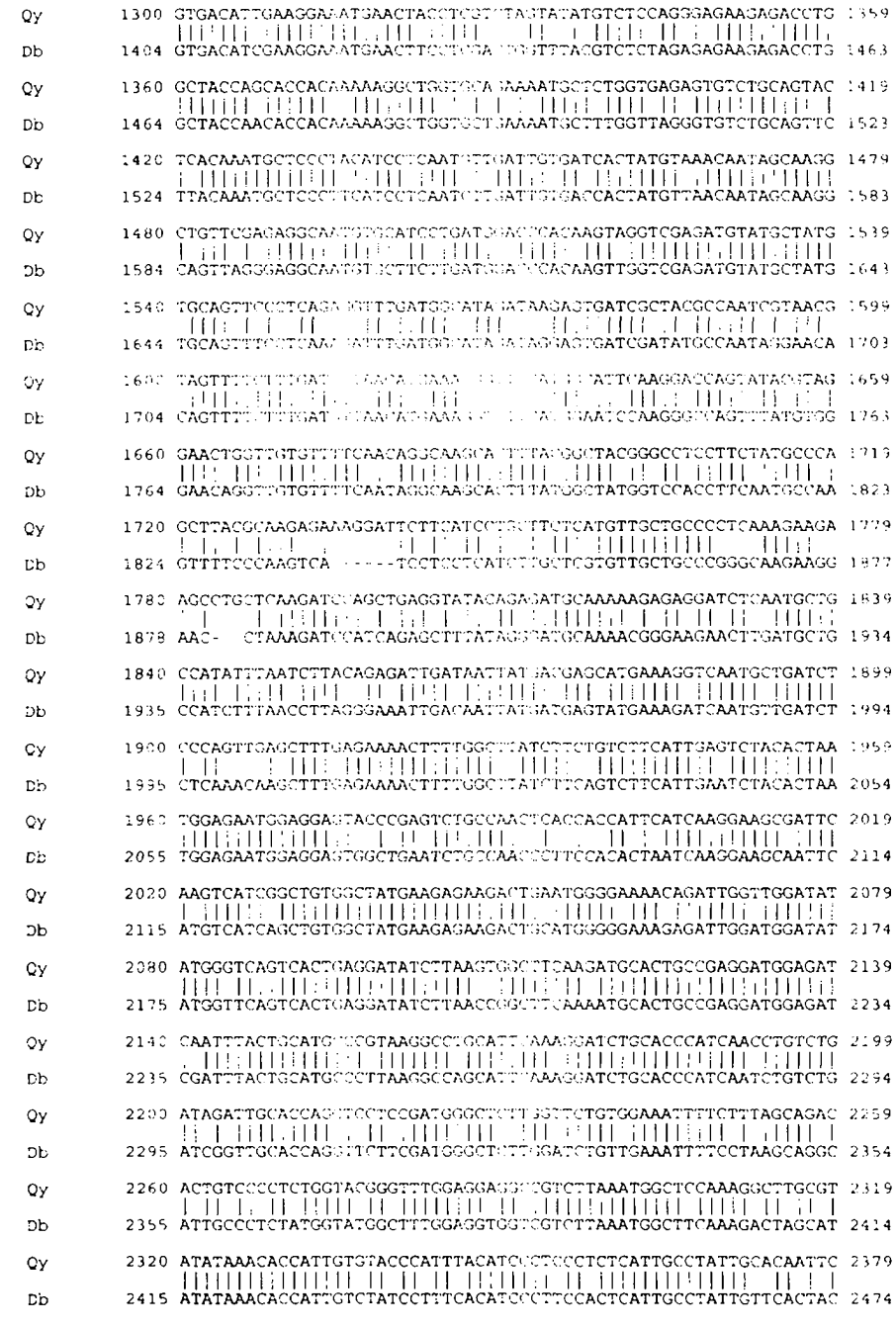
Query Match 61.6%; Score 1989.8; DB 10; Length 3328;
Best Local Similarity 80.8%; Pred. No. 0;
Matches 2399; Conservative 0; Mismatches 552; Indels 18; Gaps 6;

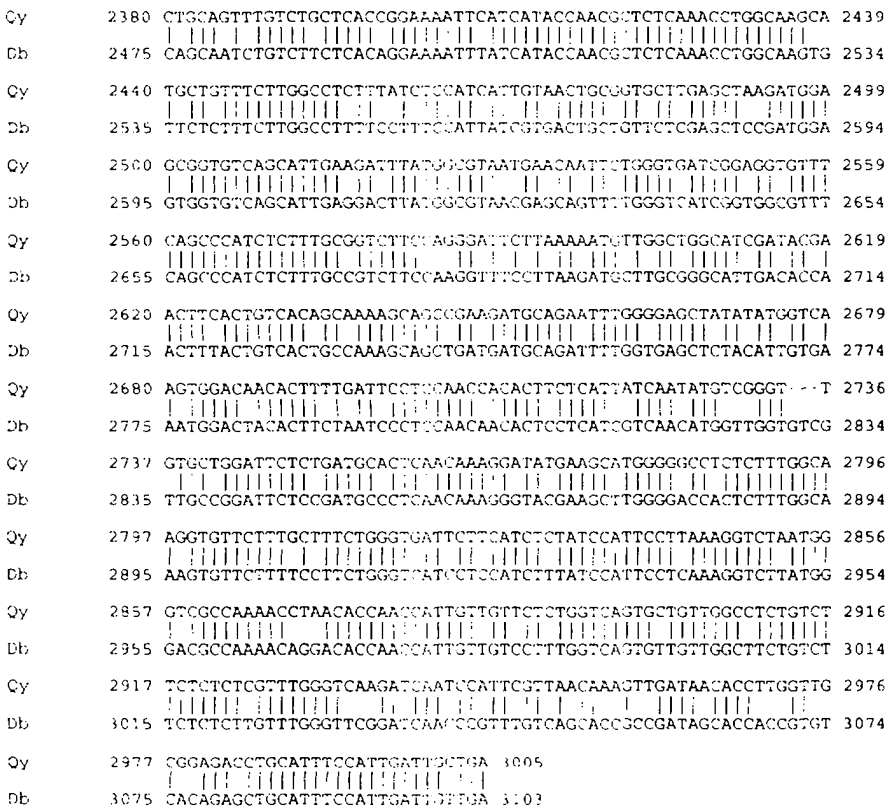
```
Qy          40 AAAGAGCTGGTAAAGTGGTAATAAGCAAGATGATGGAATCTGGGGCTCCTATATGCCATA 99
               || |||   |  ||||  |  |    | ||||||||||||||||| |||| | ||||| |
Db         150 AAGGAGAAAGATAAGTACTTTTTTTGAGAATGATGGAATCTGGGGTTCCTGTTTGCCACA 209

Qy         100 CCTGTGGTGAACAGGTGGGGCATGATGCAAATGGGGAGCTATTTGTGGCTTGCCATGAGT 159
               | |||||||||| || |||    |||  |||| || |  ||||||||||||||||| |
Db         210 CTTGTGGTGAACATGTTGGGTTGAATGTTAATGGTGAACCTTTTGTGGCTTGCCATGAAT 269

Qy         160 GTAGCTATCCCATGTGCAAGTCTTGTTTCGAGTTTGAAATCAATGAGGGCCGGAAAGTTT 219
               |||  |  | || || || |||  ||| || || || |||  | || || || || |||| ||
```

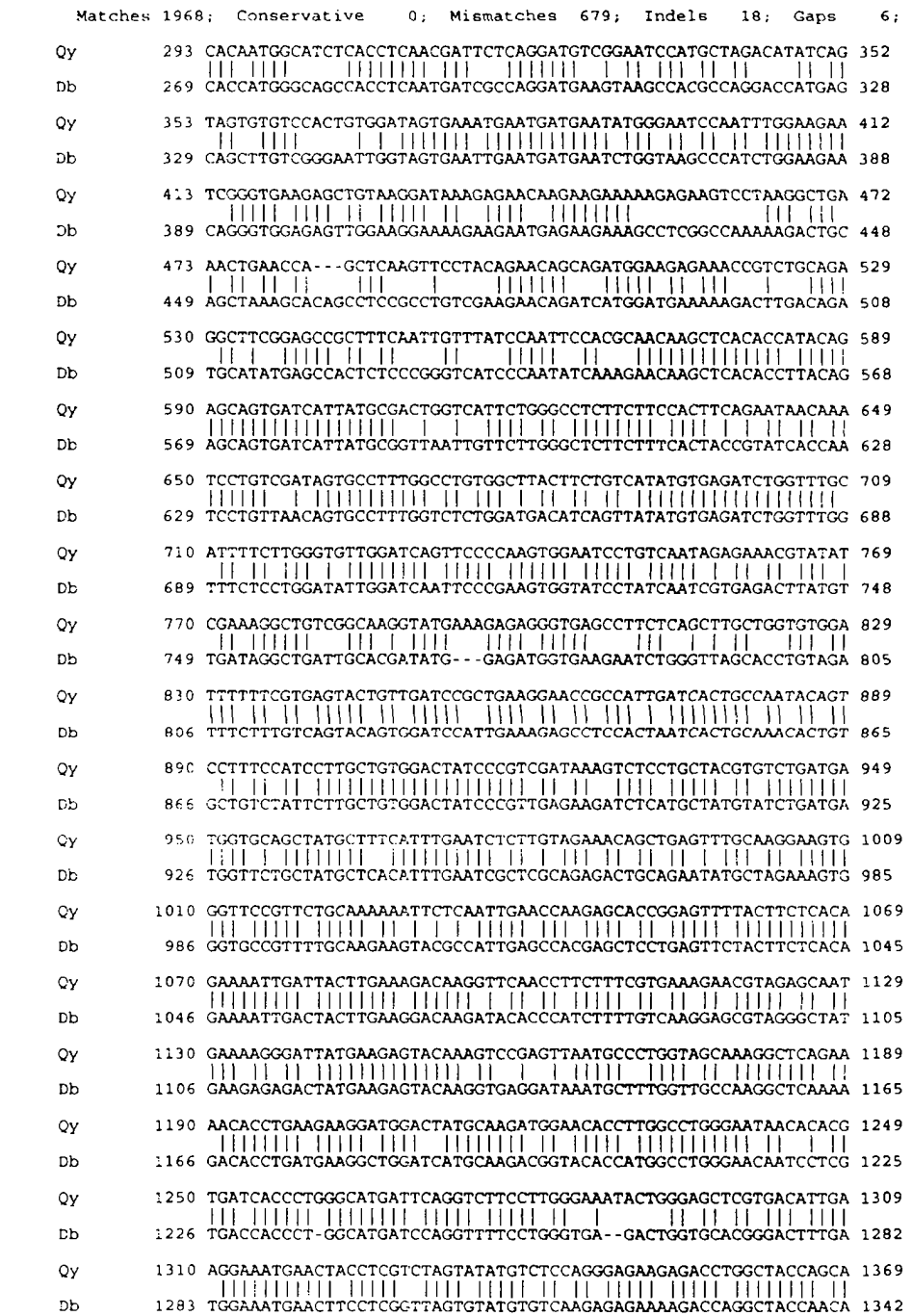

270 GTAATTTCCCTATTTGTAAGAGTTGTTTTGAGTATGATCTTAAGGAAGGACGAAAAGCTT 329
220 GCTTGCGGTGTGGCTCGCCATATGATGAGAACTTGCTGGATGATGTAGAAAAGAAGGGGT 279
330 GCTTGCGTTGTGGTAGTCCATATGATGAAAACCTGTTGGACGATGTCGA---GAAGGCCA 386
280 CTGGCAATCAATCCACAATGGCATCTCACCTCAACGATTCTCAGGATGTCGGAATCCATG 339
387 CCGGCGATCAATCGACAATGGCTGCACATTTGAACAAGTCTCAGGATGTTGGAATTCATG 446
340 CTAGACATATCAGTAGTGTGTCCACTGTGGATAGTGAAATGAATGATGAATATGGGAATC 399
447 CAAGACATATCAGCAGTGTGTCTACATTGGATAGTGAAATGGCTGAAGACAATGGGAATT 506
400 CAATTTGGAAGAATCGGGTGAAGAGCTGTAAGGATAAAGAGAACAAGAAGAAAAAGAGAA 459
507 CGATTTGGAAGAACAGGGTGGAAAGTTGGAAAGAAAAGAAGAACAAGAAGAAGAAGCCTG 566
460 GTCCTAAGGCTGAAACTGAACCAGCTCAAGTTCCTACAGAACAGCAGATGGAAGAGAAAC 519
567 CAACAACTAAGGTTGAAAGAGAGGCTGAAATCCCACCTGAGCAACAAATGGAAGATAAAC 626
520 CGTCTGCAGAGGCTTCGGAGCCGCTTTCAATTGTTTATCCAATTCCACGCAACAAGCTCA 579
627 CGGCACCGGATGCTTCCCAGCCCCTCTCGACTATAATTCCAATCCCGAAAAGCAGACTTG 686
580 CACCATACAGAGCAGTGATCATTATGCGACTGGTCATTCTGGGCCTCTTCTTCCACTTCA 639
687 CACCATACCGAACCGTGATCATTATGCGATTGATCATTCTTGGTCTTTTCTTCCATTATC 746
640 GAATAACAAATCCTGTCGATAGTGCCTTTGGCCTGTGGCTTACTTCTGTCATATGTGAGA 699
747 GAGTAACAAACCCCGTTGACAGTGCTTTTGGACTGTGGCTCACTTCAGTCATATGTGAAA 806
700 TCTGGTTTGCATTTTCTTGGGTGTTGGATCAGTTCCCCAAGTGGAATCCTGTCAATAGAG 759
807 TCTGGTTTGCATTTTCCTGGGTGTTGGATCAGTTCCCTAAGTGGTATCCTGTTAACAGGG 866
760 AAACGTATATCGAAAGGCTGTCGGCAAGGTATGAAAGAGAGGGTGAGCCTTCTCAGCTTG 819
867 AAACATACATTGACAGACTATCTGCAAGATATGAAAGAGAAGGTGAACCTGATGAACTTG 926
820 CTGGTGTGGATTTTTTCGTGAGTACTGTTGATCCGCTGAAGGAACCGCCATTGATCACTG 879
927 CTGCAGTTGACTTCTTCGTGAGTACAGTGGATCCATTGAAAGAGCCTCCATTGATTACTG 986
880 CCAATACAGTCCTTTCCATCCTTGCTGTGGACTATCCCGTCGATAAAGTCTCCTGCTACG 939
987 CCAATACTGTGCTTTCCATCCTTGCCTTGGACTACCCGGTGGATAAGGTCTCTTGTTATA 1046
940 TGTCTGATGATGGTGCAGCTATGCTTTCATTTGAATCTCTTGTAGAAACAGCTGAGTTTG 999
1047 TATCTGATGATGGTGCGGCCATGCTGACATTTGAATCTCTAGTAGAAACAGCCGACTTTG 1106
1000 CAAGGAAGTGGGTTCCGTTCTGCAAAAAATTCTCAATTGAACCAAGAGCACCGGAGTTTT 1059
1107 CAAGAAAGTGGGTTCCATTCTGCAAAAAATTTTCCATTGAACCCCGGGCACCTGAGTTTT 1166
1060 ACTTCTCACAGAAAATTGATTACTTGAAAGACAAGGTTCAACCTTCTTTCGTGAAAGAAC 1119
1167 ACTTCTCACAGAAGATTGATTACTTGAAAGATAAAGTGCAGCCCTCTTTTGTAAAAGAAC 1226
1120 GTAGAGCAATGAAAAGGGATTATGAAGAGTACAAAGTCCGAGTTAATGCCCTGGTAGCAA 1179
1227 GTAGAGCTATGAAAAGAGATTATGAAGAGTACAAAATTCGAATCAATGCTTTAGTTGCAA 1286
1180 AGGCTCAGAAAACACCTGAAGAAGGATGGACTATGCAAGATGGAACACCTTGGCCTGGGA 1239
1287 AGGCTCAGAAAACACCTGATGAAGGATGGACAATGCAAGATGGAACTTCTTGGCCAGGAA 1346
1240 ATAACACACGTGATCACCCTGGGCATGATTCAGGTCTTCCTTGGGAAATACTGGGAGCTC 1299
1347 ATAACCCGCGTGATCACCCT-GGCATGATTCAGGTTTTCCTTGG--ATATAGTGGTGCTC 1403


Qy 1300 GTGACATTGAAGGAAATGAACTACCTCGTTATATATGTCTCCAGGGAGAAGAGACCTG 1359
Db 1404 GTGACATCGAAGGAAATGAACTTCCTCGATTTACGTCTCTAGAGAGAAGAGACCTG 1463
Qy 1360 GCTACCAGCACCACAAAAAGGCTGGTGCAAAAATGCTCTGGTGAGAGTGTCTGCAGTAC 1419
Db 1464 GCTACCAACACCACAAAAAGGCTGGTGCTAAAATGCTTTGGTTAGGGTGTCTGCAGTTC 1523
Qy 1420 TCACAAATGCTCCCTACATCCTCAATCTTGATTGTGATCACTATGTAAACAATAGCAAGG 1479
Db 1524 TTACAAATGCTCCCTTCATCCTCAATCTTGATTGTGATCACTATGTTAACAATAGCAAGG 1583
Qy 1480 CTGTTCGAGAGGCAATGTGCATCCTGATGAACCACAAGTAGGTCGAGATGTATGCTATG 1539
Db 1584 CAGTTAGGGAGGCAATGTGCTTCTTGATGAACCACAAGTTGGTCGAGATGTATGCTATG 1643
Qy 1540 TGCAGTTCCTCAGA[illegible]TGATGGCATA[illegible]AAGAGTGATCGCTACGCCAATCGTAACG 1599
Db 1644 TGCAGTTTCCTCAA[illegible]TGATGGCATA[illegible]AAGAGTGATCGATATGCCAATAGGAACA 1703
Qy 1600 TAGTTTTCTTTGAT[illegible]ATTCAAGGACCAGTATACGTAG 1659
Db 1704 CAGTTTTCTTTGAT[illegible]AAGGTCCAGTTTATGTGG 1763
Qy 1660 GAACTGGTTGTGTTTTCAACAGGCAAGCA[illegible]GGCTACGGGCCTCCTTCTATGCCCA 1719
Db 1764 GAACAGGTTGTGTTTTCAATAGGCAAGCACTTTATGGCTATGGTCCACCTTCAATGCCAA 1823
Qy 1720 GCTTACGCAAGAGAAAGGATTCTTCATCCTGCTTCTCATGTTGCTGCCCCTCAAAGAAGA 1779
Db 1824 GTTTTCCCAAGTCA-----TCCTCCTCATCTTGCTCGTGTTGCTGCCCGGGCAAGAAGG 1877
Qy 1780 AGCCTGCTCAAGATCAAGCTGAGGTATACAGAGATGCAAAAAGAGAGGATCTCAATGCTG 1839
Db 1878 AAC-CTAAAGATCCATCAGAGCTTTATAGGTATGCAAAACGGGAAGAACTTGATGCTG 1934
Qy 1840 CCATATTTAATCTTACAGAGATTGATAATTATGACGAGCATGAAAGGTCAATGCTGATCT 1899
Db 1935 CCATCTTTAACCTTAGGGAAATTGACAATTATGATGAGTATGAAAGATCAATGTTGATCT 1994
Qy 1900 CCCAGTTGAGCTTTGAGAAAACTTTTGGCTATCTTCTGTCTTCATTGAGTCTACACTAA 1959
Db 1935 CTCAAACAAGCTTTGAGAAAACTTTTGGCTTATCTTCAGTCTTCATTGAATCTACACTAA 2054
Qy 1960 TGGAGAATGGAGGAGTACCCGAGTCTGCCAACTCACCACCATTCATCAAGGAAGCGATTC 2019
Db 2055 TGGAGAATGGAGGAGTGGCTGAATCTGCCAACTCTTCCACACTAATCAAGGAAGCAATTC 2114
Qy 2020 AAGTCATCGGCTGTGGCTATGAAGAGAAGACTGAATGGGGAAAACAGATTGGTTGGATAT 2079
Db 2115 ATGTCATCAGCTGTGGCTATGAAGAGAAGACTGCATGGGGGAAAGAGATTGGATGGATAT 2174
Qy 2080 ATGGGTCAGTCACTGAGGATATCTTAAGTGGCTTCAAGATGCACTGCCGAGGATGGAGAT 2139
Db 2175 ATGGTTCAGTCACTGAGGATATCTTAACCGGCTTCAAAATGCACTGCCGAGGATGGAGAT 2234
Qy 2140 CAATTTACTGCATG[illegible]GTAAGGCCTGCATT[illegible]AAAGGATCTGCACCCATCAACCTGTCTG 2199
Db 2235 CGATTTACTGCATGCGCTTAAGGCCAGCATT[illegible]AAAGGATCTGCACCCATCAATCTGTCTG 2294
Qy 2200 ATAGATTGCACCAG[illegible]CCTCCGATGGGCTCTT[illegible]TCTGTGGAAATTTTCTTTAGCAGAC 2259
Db 2295 ATCGGTTGCACCAG[illegible]TCTTCGATGGGCTCTT[illegible]TCTGTTGAAATTTTCCTAAGCAGGC 2354
Qy 2260 ACTGTCCCCTCTGGTACGGGTTTGGAGGA[illegible]GTCTTAAATGGCTCCAAAGGCTTGCGT 2319
Db 2355 ATTGCCCTCTATGGTATGGCTTTGGAGGTGGTCGTCTTAAATGGCTTCAAAGACTAGCAT 2414
Qy 2320 ATATAAACACCATTGTGTACCCATTTACATC[illegible]TCCTCTCATTGCCTATTGCACAATTC 2379
Db 2415 ATATAAACACCATTGTCTATCCTTTCACATC[illegible]TCCACTCATTGCCTATTGTTCACTAC 2474



RESULT 2
US 10 209-059 27
; Sequence 27, Application US/10...
; Publication No. US20030169838A.
; GENERAL INFORMATION:
; APPLICANT: Dhugga, Kanwarpal S.
; APPLICANT: Wang, Haiyin
; TITLE OF INVENTION: Maize Cellulose Synthases and Uses
; TITLE OF INVENTION: Thereof
; FILE REFERENCE: 0864R2
; CURRENT APPLICATION NUMBER: US/10/209,059
; CURRENT FILING DATE: 2002-07-31
; PRIOR APPLICATION NUMBER: 60/096,812
; PRIOR FILING DATE: 1998-08-17
; PRIOR APPLICATION NUMBER: 09/371,383
; PRIOR FILING DATE: 1999-08-06
; PRIOR APPLICATION NUMBER: 09/550,483
; PRIOR FILING DATE: 2000-04-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 3231
; TYPE: DNA
; ORGANISM: Zea mays
US-10-209-059-27

Query Match 45.9%; Score 1482.6; DB 12; Length 3231;
Best Local Similarity 73.8%; Pred. No. 0;



```
1370 CCACAAAAAGGCTGGTGCAGAAAATGCTCTGGTGAGAGTGTCTGCAGTACTCACAAATGC 1429
1343 CCACAAGAAGGCAGGGGCTATGAATGCTCTGGTCCGAGTGTCTGCTGTTCTGACAAATGC 1402

1430 TCCCTACATCCTCAATGTTGATTGTGATCACTATGTAAACAATAGCAAGGCTGTTCGAGA 1489
1403 CCCTTACATTCTTAATCTTGATTGTGATCACTATGTTAACAACAGCAAAGCTGTTCGTGA 1462

1490 GGCAATGTGCATCCTGATGGACCCACAAGTAGGTCGAGATGTATGCTATGTGCAGTTCCC 1549
1463 AGCAATGTGCTTCATGATGGACCCTACTGTTGGCAGAGATGTCTGCTATGTACAATTCCC 1522

1550 TCAGAGGTTTGATGGCATAGATAAGAGTGATCGCTACGCCAATCGTAACGTAGTTTTCTT 1609
1523 CCAGAGGTTCGATGGCATTGATCGCAGTGATCGATATGCCAATAGGAACGTTGTGTTCTT 1582

1610 TGATGTTAACATGAAAGGGTTGGATGGCATTCAAGGACCAGTATACGTAGGAACTGGTTG 1669
1583 TGATGTTAATATGAAAGGACTTGATGGCCTCCAAGGCCCAGTTTATGTGGGAACTGGTTG 1642

1670 TGTTTTCAACAGGCAAGCACTTTACGGCTACGGGCCTCCTTCTATGCCCAGTTACGCAA 1729
1643 TTGTTTCAATAGGCAAGCACTTTATGGTTATGGGCCTCCATCTCTGCCCGCACTTCCAAA 1702

1730 GAGAAAGGATTCTTCATCCTGCTTCTCATGTTGCTGCCCCTCAAAGAAGAAGCCTGCTCA 1789
1703 GTCTTCGATTTGTTC------CTGGTGTTGCTGCTGCTGTCCCAAGAAAAAGGTTGAAAG 1756

1790 AGATCCAGCTGAGGTATACAGAGATGCAAAAAGAGAGGATCTCAATGCTGCCATATTTAA 1849
1757 AAGTGAGAGGGAAATCAACAGAGACTCTCGGCGAGAAGACCTCGAGTCTGCCATTTTTAA 1816

1850 TCTTACAGAGATTGATAATTATGACGAGCATGAAAGGTCAATGCTGATCTCCCAGTTGAG 1909
1817 CCTTCGCGAAATTGACAACTACGATGAGTACGAGAGGTCCATGCTCATCTCTCAGATGAG 1876

1910 CTTTGAGAAAACTTTTGGCTTATCTTCTGTCTTCATTGAGTCTACACTAATGGAGAATGG 1969
1877 CTTCGAGAAGTCTTTTGGGCTGTCCTCGGTCTTTATTGAATCGACCCTTATGGAGAATGG 1936

1970 AGGAGTACCCGAGTCTGCCAACTCACCACCATTCATCAAGGAAGCGATTCAAGTCATCGG 2029
1937 GGGCGTCCCTGAATCTGCAAACCCATCTACCCTAATTAAAGAAGCCATTCATGTCATTAG 1996

2030 CTGTGGCTATGAAGAGAAGACTGAATGGGGAAAACAGATTGGTTGGATATATGGGTCAGT 2089
1997 CTGTGGATATGAAGAGAAAACTGAATGGGGAAAAGAGATTGGCTGGATCTATGGTTCAGT 2056

2090 CACTGAGGATATCTTAAGTGGCTTCAAGATGCACTGCCGAGGATGGAGATCAATTTACTG 2149
2057 TACAGAGGATATTCTGACTGGGTTTAAGATGCACTGCCGTGGCTGGAGATCCATCTACTG 2116

2150 CATGCCCGTAAGGCCTGCATTCAAAGGATCTGCACCCATCAACCTGTCTGATAGATTGCA 2209
2117 CATGCCGGTGAGACCTGCATTCAAGGGATCAGCCCCAATCAATCTTTCCGATCGTCTTCA 2176

2210 CCAGGTCCTCCGATGGGCTCTTGGTTCTGTGGAAATTTTCTTTAGCAGACACTGTCCCCT 2269
2177 CCAAGTTCTCCGGTGGGCTCTTGTTTCTGTCGAGATCTTCTTCAGTCGGCACTGCCCGCT 2236

2270 CTGGTACGGGTTTGGAGGAGGCCGTCTTAAATGGCTCCAAAGGCTTGCGTATATAAACAC 2329
2237 GTGGTACGGTTACGGTGGCGGCCGTCTGAAATGGCTCCAGAGGCTCTCCTACATCAACAC 2296

2330 CATTGTGTACCCATTTACATCCCTCCCTCTCATTGCCTATTGCACAATTCCTGCAGTTTG 2389
2297 CATCGTGTACCCGTTCACTTCTCTTCCTCTCGTTGCCTACTGTTGCCTGCCTGCCATTTG 2356

2390 TCTGCTCACCGGAAAATTCATCATACCAACGCTCTCAAACCTGGCAAGCATGCTGTTTCT 2449
2357 CCTGCTCACAGGAAAGTTCATTATACCTACGCTGTCCAACGCTGCAACGATATGGTTTCT 2416
```

```
Qy      2450 TGGCCTCTTTATCTCCATCATTGTAACTGCGGTGCTTGAGCTAAGATGGAGCGGTGTCAG 2509
Db      2417 TGGCCTCTTCATGTCCATCATCGTGACGACGTGTTGGAGCTGCGGTGGAGTGGCATCGG 2476

Qy      2510 CATTGAAGATTTATGCGTAATGAACAATTCTGGTGATCGGAGGTGTTTCAGCCCATCT 2569
Db      2477 GATCGAGGACTGGTGGCGCAACGAGGAGTTCTGGTCATCGGAGGCGTGTCCGCGCACCT 2536

Qy      2570 CTTTGCGGTCTTCCAAGGATTCTTAAAAATGTTGGCTGGCATCGATACGAACTTCACTGT 2629
Db      2537 GTTCGCCGTGTTCCAAGGTATCCTCAAAATGATTGCCGGGCTGGACACCAACTTCACGGT 2596

Qy      2630 CACAGCAAAAGCAGCTGAAGATGCAGAATTTGAGGAGCTATATATGGTCAAGTGGACAAC 2689
Db      2597 CACGGCAAAGGCCAAGGACGACACTGAGTTCGAGGAGCTGTACCTGTTCAAGTGGACGAC 2656

Qy      2690 ACTTTTGATTCCTCCAACCACACTTCTCATTATCAATATGTCGGG---TTGTGCTGGATT 2746
Db      2657 GGTGCTGATCCCGC[illegible]AAGCATCT[illegible]GAACCTGGTGGGCGTGGTGGCTGGGTT 2716

Qy      2747 CTCTGATGCACTCA[illegible]AAAGGATA[illegible]GGCGTCTCTTTGGCAAGGTGTTCTT 2806
Db      2717 CTCGGC[illegible]GCTTCAA[illegible]GCGCTCGTTCGGTAAGGTGTTCTT 2776

Qy      2807 TGCTTTCTGGGTGATCCTTCATCTCTATCATCCTTAAAGGTCTAATGGGTCGCCAAAA 2866
Db      2777 CGCCATGTGGGTGATCATGCACCTGTACCGGTTCTCAAGGGTCTCATGGGCCGCCAGAA 2836

Qy      2867 CCTAACACCAACCATTGTTGTTCTCTGGTCAGTGCTGTTGGCCTCTGTCTTCTCTCTCGT 2926
Db      2837 CCGCACGCCGACCATCGTGGTGCTCTGGTCCTTGCTCCTCGCCTCCGTCTTCTCCCTCCT 2896

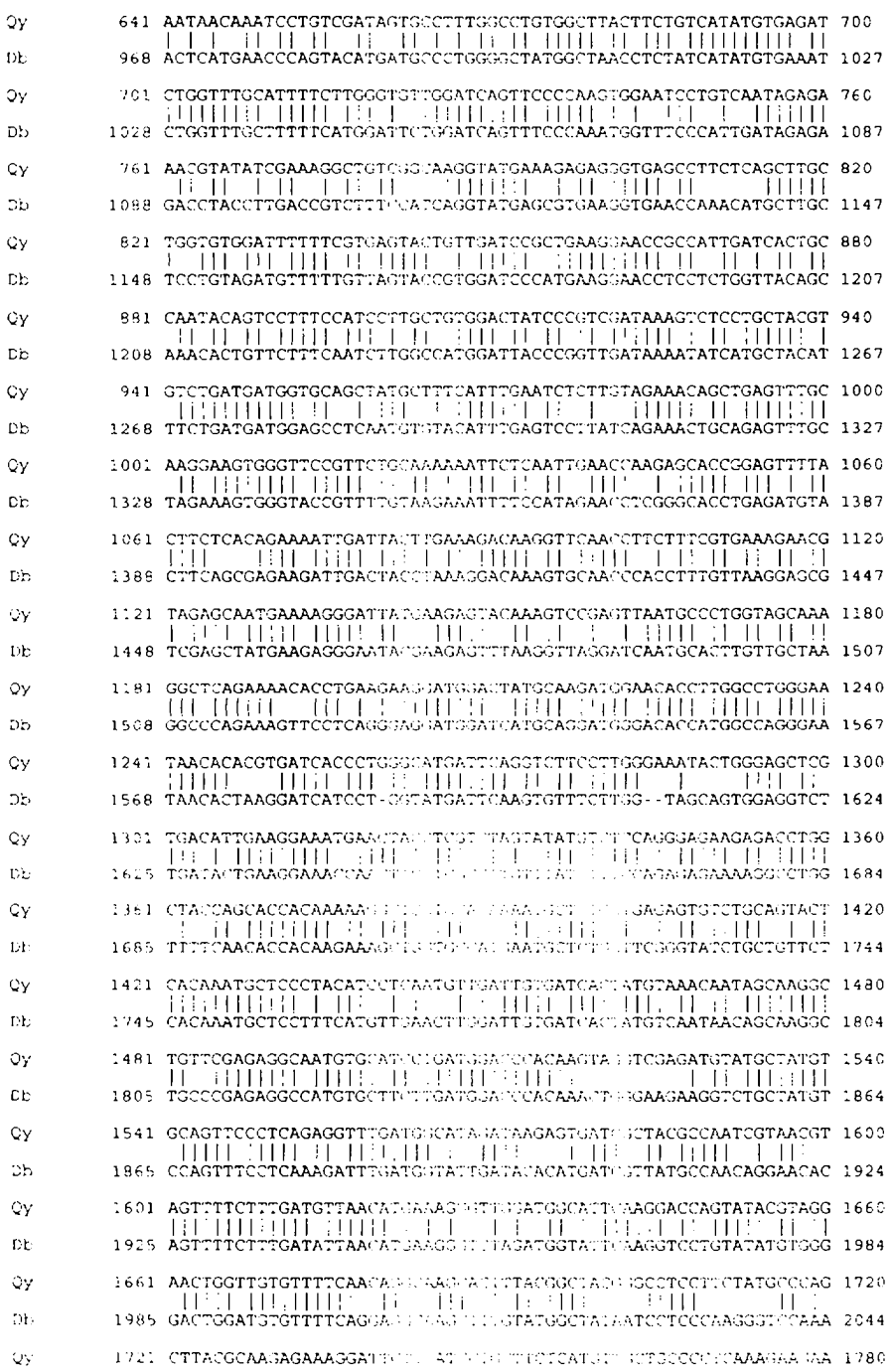
Qy      2927 TTGGGTCAAGATCAATCCATTCGTT 2951
Db      2897 GTGGGTCAAGATCGACCCATTCGTT 2921
```

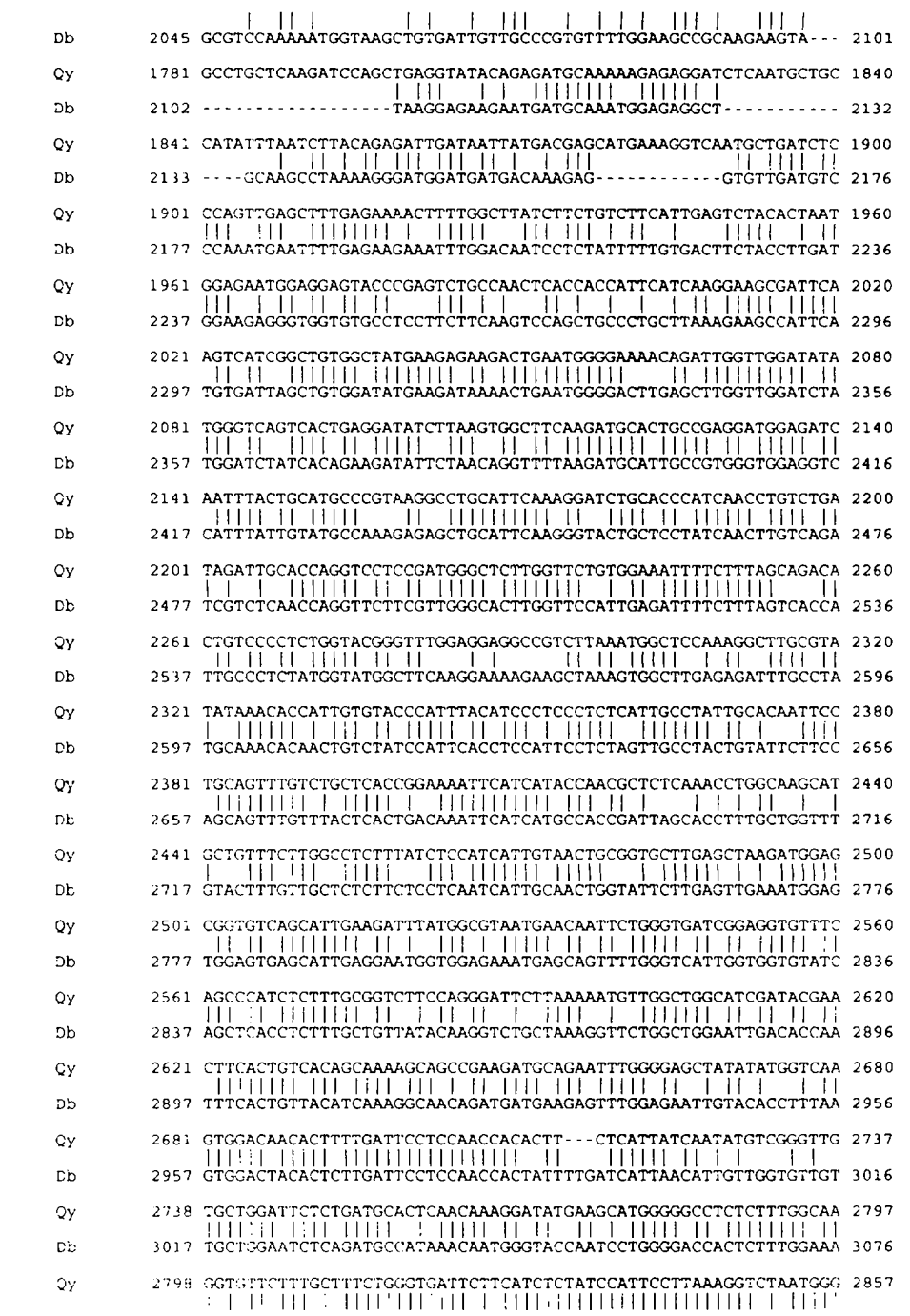
```
RESULT 3
US-09-900-237-13
; Sequence 13, Application US/09900237
; Patent No. US20020124A1
; GENERAL INFORMATION:
;  APPLICANT: Allen, Stephen
;  TITLE OF INVENTION: Plant Cellulose Synthases
;  FILE REFERENCE: BB1170 US CIP
;  CURRENT APPLICATION NUMBER: US/09/900,237
;  CURRENT FILING DATE:  2001-07-06
;  PRIOR APPLICATION NUMBER: 60/092,844
;  PRIOR FILING DATE: 1998-07-14
;  PRIOR APPLICATION NUMBER: PCT/US99/15871
;  PRIOR FILING DATE: 1999-07-13
;  PRIOR APPLICATION NUMBER: 09/720383
;  PRIOR FILING DATE: 2000-12-21
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 13
;   LENGTH: 3517
;   TYPE: DNA
;   ORGANISM: Glycine max
US-09-900-237-13

  Query Match           35.7%;  Score 1152.6;  DB 10;  Length 3517;
  Best Local Similarity   69.1%;  Pred. No. 0;
  Matches 1686;  Conservative    0;  Mismatches  699;  Indels   54;  Gaps    6;

Qy        521 GTCTGCAGAGGCTTCGGAGCCGCTTTCAATTGTTTATCCAATTCCACGCAACAAGCTCAC 580
Db        848 GTTAGATGAAGCAAGGCAACCACTGTCAAGGAAAGTGCCAATAGCATCCAGCAAAATCAA 907

Qy        581 ACCATACAGAGCAGTGATCATTATGCGACTGGTCATTCTGGGCCTCTTCTTCCACTTCAG 640
Db        908 TCCATATAGAATGGTGATTGTGGCACGTCTGGTTATTCTTGCTTTCTTCCTCAGATACAG 967
```





```
      3077 GCTCTTCTTTTCCTTCTGGGTGATTGTCCATCTCTATCCATTCCTTAAAGGTTTGATGGG 3136

      2858 TCGCCAAAACCTAACACCAACCATTGTTGTTCTCTGGTCAGTGCTGTTGGCCTCTGTCTT 2917
           ||||||||| | ||||| ||||||||||  | ||||||||||| |||||||| | ||
      3137 TCGCCAAAATCGCACACCCACCATTGTTGTGATTTGGTCAGTGCTATTGGCCTCTATTTT 3196

      2918 CTCTCTCGTTTGGGTCAAGATCAATCCATTCGTTAACAA 2956
           |||  |  |||||||  |  ||  |||||||  ||   |||
      3197 CTCCTTACTTTGGGTAAGAATTGATCCATTTGTCCTCAA 3235


SULT 4
-10-160-719-33
Sequence 33, Application US/10160719
Publication No. US20030167528A1
GENERAL INFORMATION:
 APPLICANT: Kanwarpal S. Dhugga
 APPLICANT:  Timothy G. Helentjaris
 APPLICANT:  Benjamin A.  Bowen
 APPLICANT:  Xun Wang
 TITLE OF INVENTION: Maize Cellulose Synthases and Uses
 TITLE OF INVENTION:  Thereof
 FILE REFERENCE: 0864
 CURRENT APPLICATION NUMBER: US/10/160,719
 CURRENT FILING DATE:  2002-06-03
 PRIOR APPLICATION NUMBER: US/09/371,383A
 PRIOR FILING DATE: 1999-08-06
 PRIOR APPLICATION NUMBER: US 60/096,822
 PRIOR FILING DATE: 1998-08-17
 NUMBER OF SEQ ID NOS: 60
 SOFTWARE: FastSEQ for Windows Version 3.0
SEQ ID NO 33
  LENGTH: 3753
  TYPE: DNA
  ORGANISM: Zea mays
  FEATURE:
  NAME/KEY: CDS
  LOCATION: (184)...(3408)
3-10-160-719-33

Query Match             35.6%;  Score 1149.8;  DB 12;  Length 3753;
Best Local Similarity   67.1%;  Pred. No. 0;
Matches 1735;  Conservative    0;  Mismatches  836;  Indels   15;  Gaps    7;

Qy        374 TGAAATGAATGATGAATATGGGAATCCAATTTGGAAGAATCGGGTGAAGAGCTGTAAGGA 433
              ||  | |||  ||    |  ||||   ||      ||  |   |     |  |
Db        780 TGGGCTTAATAGTGTTGACTGGAAGGAAAGAGTTGAGAGCTGGAGGGTTAAACAGGACAA 839

Qy        434 TAAAGAGAACAAGAAGAAAAAAGAGAAGTCCTAAGGCTGAAACTGAACCAGCTCAAGTTCC 493
              ||  ||   |  ||  || | |  |||  |||||  |   | |        | |   |
Db        840 AAATATGATGCAAGTGACTAATAAATATCCAGAGGCTAGAGGAGGAGACATGGAGGGGAC 899

Qy        494 TACAGAACAGCAGATGGAAGAGAAACCGTCTGCAGAGGCTTCGGAGCCGCTTTCAATTGT 553
              |     | |       ||   | ||  |  ||  ||  ||    |  ||  |      | |
Db        900 TGGCTCAAATGGAGAAGATATGCAAATGGTTGATGATGCACGGCTACCTTTGAGCCGTAT 959

Qy        554 TTATCCAATTCCACGCAACAAGCTCACACCATACAGAGCAGTGATCATTATGCGACTGGT 613
              ||||||  |    ||| ||||||  |  ||| | | |||||||||| | ||  |
Db        960 CGTGCCAATTTCCTCAAACCAGCTCAACCTTTACCGGGTAGTGATCATTCTCCGTCTTAT 1019

Qy        614 CATTCTGGGCCTCTTCTTCCACTTCAGAATAACAAATCCTGTCGATAGTGCCTTTGGCCT 673
              ||| ||| || ||||||||||  |   |  |  |    |||  |  |  | |||  |
Db       1020 CATCCTGTGCTTCTTCTTCCAGTATCGTGTCAGTCATCCAGTGCGTGATGCTTATGGATT 1079

Qy        674 GTGGCTTACTTCTGTCATATGTGAGATCTGGTTTGCATTTTCTTGGGTGTTGGATCAGTT 733
              |||||    |||||  || || ||| ||||||||||  || ||||||  |  | ||||||||
Db       1080 ATGGCTAGTATCTGTTATCTGCGAGGTCTGGTTTGCCTTGTCTTGGCTTCTAGATCAGTT 1139

Qy        734 CCCCAAGTGGAATCCTGTCAATAGAGAAACGTATATCGAAAGGCTGTCGGCAAGGTATGA 793
              ||| || ||| ||||  ||||  | || || ||| | || |||||  |    ||||||||
Db       1140 CCCAAAATGGTATCCAATCAACCGTGAGACATATCTTGACAGGCTTGCATTGAGGTATGA 1199

Qy        794 AAGAGAGGGTGAGCCTTCTCAGCTTGCTGGTGTGGATTTTTTCGTGAGTACTGTTGATCC 853
```

```
Db        1200 TAGAGAGGGAGAGCTAGTACAGTTGTCTGTGATTGATGTCTTCGTCAGTACAGTGGATCC 1259

Qy         854 GCTGAAGGAACCGGCAATTGATCACTTAATAAGTCTTTTCCATCCTTGCTGTGGACTA 913
               |||||||||| ||  |||  ||  ||  ||  | |||| ||| |||||| || ||
Db        1260 ATTGAAGGAACCTCCACTGATCACAGTAACATGTTTTGTCCATTCTTTCTGTGGATTA 1319

Qy         914 TCCCGTCGATAAAGTTTCCTGCTACGAGCTTGATGATGGTGCAGCTATGCTTTCATTTGA 973
               ||  || ||  |||  | |||  ||  ||  |||| ||||||||||  | ||||
Db        1320 CCCTGTTGACAAAGTGTCATGCTATGTTTCTGATGATGGTTCAGCTATGCTGACTTTTGA 1379

Qy         974 ATCTCTTGTAGAAAAAGCTGAGTTTGCAAAGAAGTGGGTTCCGTTCTGCAAAAAATTCTC 1033
               ||||  |||||  | || ||  ||| |||||| || || || ||  |
Db        1380 GTCTCTCTCAGAAACTGAGAATTTGCTAGAAAGTGGGTTCCCTTTTGTAACAAGCACAA 1439

Qy        1034 AATTGAACCAAGAGCAGTGGAGTTTTACTTTCTCACAGAAAATTGATTACTTCAAAGACAA 1093
               ||||||||||||||  || |  || ||  || |||| |||||  ||| ||||
Db        1440 TATTGAACCAAGAGTTGAGAATTTTACTTTGTTCAAAAAATAGATTACCTGAAGGACAA 1499

Qy        1094 GGTTCAACCTTCTTTGAAAAAAGTTCTAAGAAGTAAAAGGGATTATGAAGAGTACAA 1153
               |||||||||  |  | ||  |  |  |||| ||||  |||||  | | ||
Db        1500 AATTCAAGCTTCATGGAAAGAAAGTTCTAGAGGGAGTATGAAGAATTCAA 1559

Qy        1154 AGTCCGAGTTAATGTTCTGTAGTAAAAAGTCTTGAAAACACCTGAAGAAGGATGGACTAT 1213
               |  |  | |||| | || | | | | ||  ||||||  || | ||| ||
Db        1560 AGTAAGAATCAATGCCCTTGTTGCCAAAGTACAGAAAGTGCCTGAAGAGGGGTGGACCAT 1619

Qy        1214 GCAAGATGGAACACCTTGGCCTGGGAATAACACAGTGATCACCCTGGGCATGATTCAGG 1273
               |   |||||  | |||||| ||  |   || || |||  |||| |||  |||
Db        1620 GGCTGATGGAACTGCATGCCTGGGAATAATGTAGGGACCATCCT-GGCATGATTCAGG 1678

Qy        1274 TCTTCCTTGGGAAATACTGGGAGCTCGTGACATGAAGGAAATGAACTACCTCGTCTAGT 1333
               | | | |||  |     | ||  |  | |||||||  |||| ||  ||
Db        1679 TTTTCTTGGGGCACAGTGGTGGGCTTCTGATATTGATGGAAATGAGTTACCACGTCTTGT 1738

Qy        1334 ATATGTCTCCAGGGAAAGAGACCTGGTTACAGCACCACAAAAAGGCTGGTGCAGAAAA 1393
               |||  ||||  | |  |||  |||  |||  ||||| |||||||||||  ||
Db        1739 CTATGTCTCTCGTGAAAAGAGACCAGGCTTTTAGCATCACAAGAAGGCTGGTGCAATGAA 1798

Qy        1394 TGCTCTGGTGAGAGTGTCTGCAGTACTCACAAATGCTCCCTACATCCTCAATGTTGATTG 1453
               .||  ||  | |  ||||| ||  | |||  ||  | |||  | ||||  || ||||
Db        1799 TGCGCTGATTCGTGTATCTGCTGTGCTGACAAATGGTGCCTATCTTCTCAATGTGGATTG 1858

Qy        1454 TGATCACTATGTAAACAATAGCAAGGCTGTTCTAGAGGCAATGTGCATCCTGATGGACCC 1513
               || || ||  | ||  | ||| | | ||  | |||||||  | ||||| ||
Db        1859 CGACCATTACTTCAATAGCAGCAAAGCTCTTAGAGAAGCAATGTGCTTCATGATGGATCC 1918

Qy        1514 ACAAGTAGGTCGAGATGTATGCTATGTGCAGTTCCCTCAGAGGTTTGATGGCATAGATAA 1573
               |  |  |  |  || | |  |||| |||||  ||||||||||| ||
Db        1919 GGCTCTAGGAAGGAAAACTTGTTATGTACAATTTCCACAGAGATTTGATGGCATTGACTT 1978

Qy        1574 GAGTGATCGCTACGGCAATCGTAACGTAGTTTTCTTTGATGTTAACATGAAAGGGTTGGA 1633
               |  ||||| || || |||| | | ||| |||| |||| | ||||||||||  |||
Db        1979 GCACGATCGATATGCTAATCGGAACATAGTTTTCTTTGATATCAACATGAAAGGTCTGGA 2038

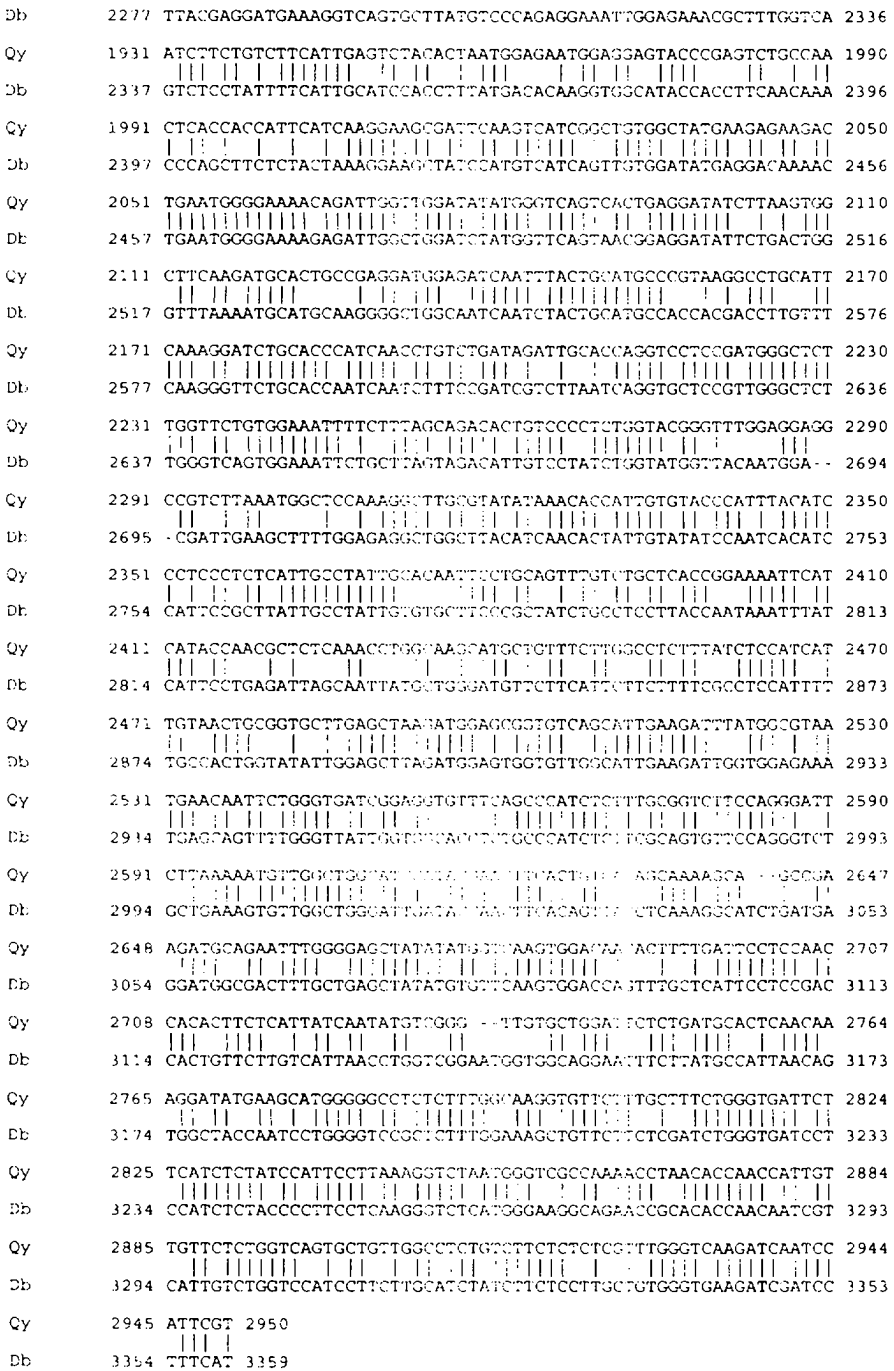
Qy        1634 TGGCATTCAAGGACCAGTATACGTACGAACTGTTTGTTTTCAACAGGCAAGCACTTTA 1693
               |||||||| || ||| ||||  |||  || |  |||||| || ||  | ||
Db        2039 TGGCATTCAGGGTCCAGTTTACGTGGGAACAGTATGCTGTTTCAATAGACAGGCTTTGTA 2098

Qy        1694 CGGCTACGGGCCTCCTTCTATGCCCAGCTTACTCAAGAGAAAGGATTCTTCATCCTGCTT 1753
               || |||| |||  |  |  |  | || |  |||  |  |  |  |||  |||
Db        2099 TGGATACGATCCTGTTTTGACTGAAGCTGATCTTGAGCCAAACATTGTTATTAAGAGCTG 2158

Qy        1754 CTCATGTTGCTGCCTTCAAAGAAGAAGCCTGTTCAAGAT-CCAGCTGAGGTATACAGAG 1812
               ||  | | |  |  |||| |  |||  |  |||  |  ||| |||  ||
Db        2159 CTGTGGTAGAAGGAAGAAAAAGAACAAGAGTTATATGGATAGTCAAAGCCGTATTATGAA 2218

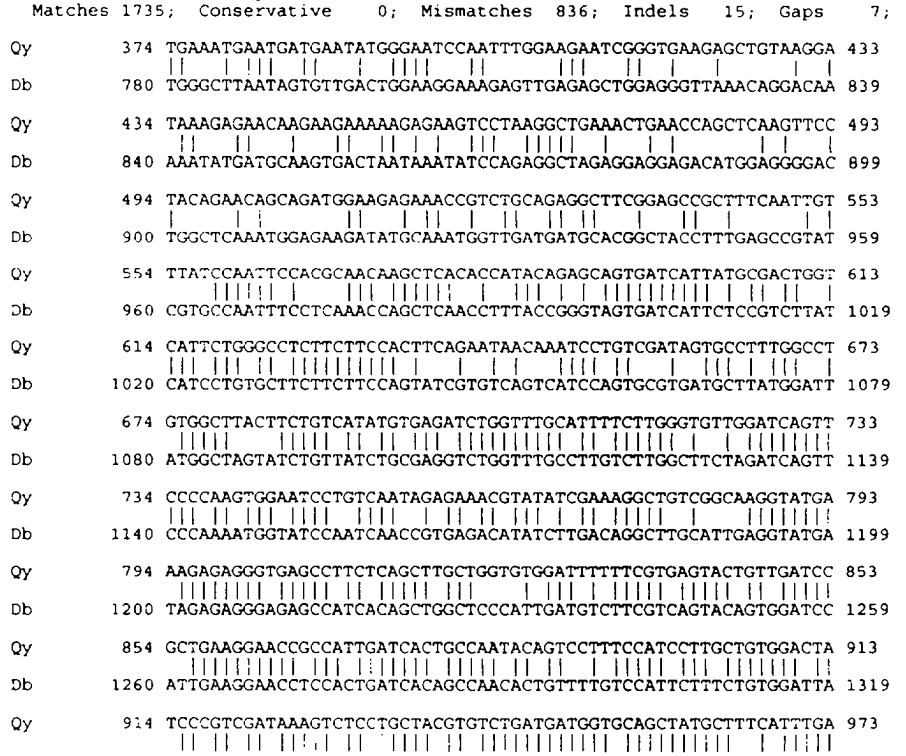
Qy        1813 ATGCAAAAAAGAGAGGATCTCAATGCTGCCATATTTA-ATCTTACAGAGATTGATAATTA 1870
               | | | |  |||  | | | | |  |  |  ||| |  |  ||
Db        2219 GAGAACAGAATCTTCAGCTCCCATCTTCAATATGGAAGACATCGAAGAGGGTATTGAAGG 2278

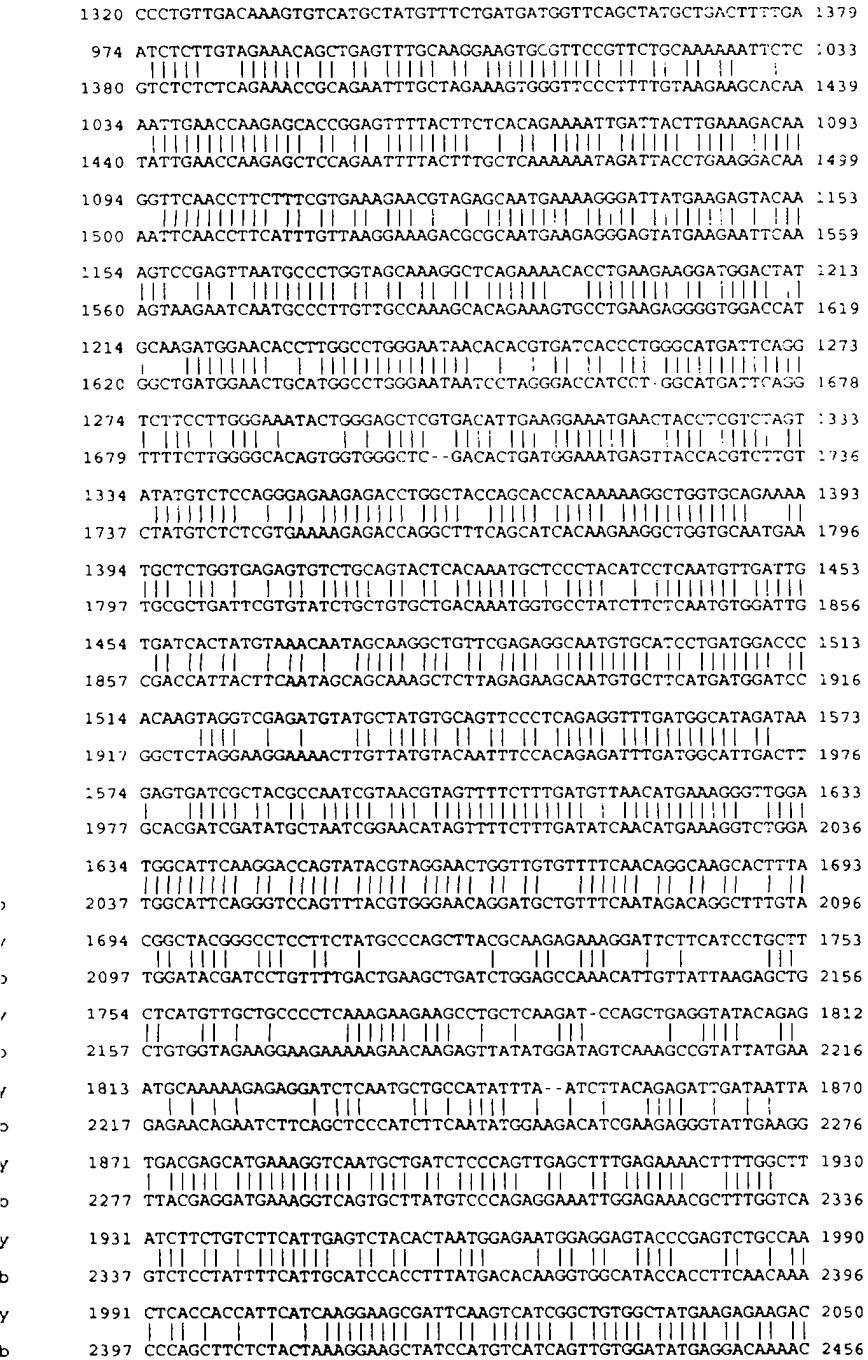
Qy        1871 TGACGAGCATGAAAGGTCAATGCTGATCTCCCAGTTGAGCTTTGAGAAAACTTTTGGCTT 1930
               | |||||  |||||||:|| | ||  | |||  |  |  || |:||||   ||||
```

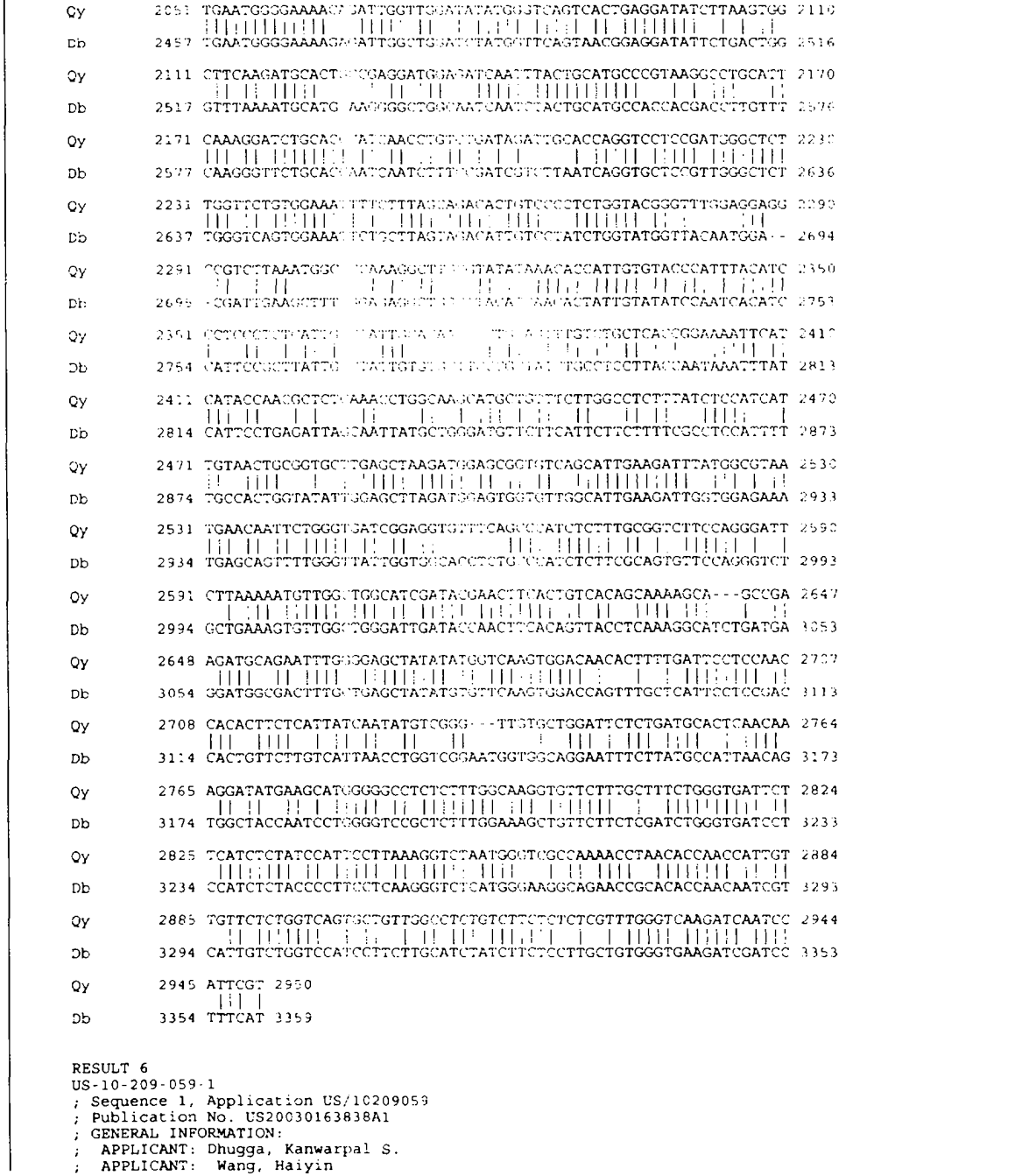


RESULT 5
US-10-160-719-53
; Sequence 53, Application US/10160719
; Publication No. US20030167528A1
; GENERAL INFORMATION:
; APPLICANT: Kanwarpal S. Dhugga
; APPLICANT: Timothy G. Helentjaris
; APPLICANT: Benjamin A. Bowen
; APPLICANT: Xun Wang
; TITLE OF INVENTION: Maize Cellulose Synthases and Uses
; TITLE OF INVENTION: Thereof
; FILE REFERENCE: 0864
; CURRENT APPLICATION NUMBER: US/10/160,719
; CURRENT FILING DATE: 2002-06-03
; PRIOR APPLICATION NUMBER: US/09/371,383A
; PRIOR FILING DATE: 1999-08-06
; PRIOR APPLICATION NUMBER: US 60/096,822
; PRIOR FILING DATE: 1998-08-17
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 53
; LENGTH: 3753
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (184)...(3408)
US-10-160-719-53

Query Match 35.6%; Score 1149.8; DB 12; Length 3753;
Best Local Similarity 67.1%; Pred. No. 0;
Matches 1735; Conservative 0; Mismatches 836; Indels 15; Gaps 7;







RESULT 6
US-10-209-059-1
; Sequence 1, Application US/10209059
; Publication No. US20030163838A1
; GENERAL INFORMATION:
; APPLICANT: Dhugga, Kanwarpal S.
; APPLICANT: Wang, Haiyin

```
;  TITLE OF INVENTION: Maize Cellulose Synthases and Uses
;  TITLE OF INVENTION:  Thereof
;  FILE REFERENCE: 0864R2
;  CURRENT APPLICATION NUMBER: US/10/209,059
;  CURRENT FILING DATE:  2002-07-31
;  PRIOR APPLICATION NUMBER: 60/096,822
;  PRIOR FILING DATE: 1998-08-17
;  PRIOR APPLICATION NUMBER: 09/371,283
;  PRIOR FILING DATE: 1999-08-06
;  PRIOR APPLICATION NUMBER: 09/550,483
;  PRIOR FILING DATE: 2000-04-14
;  NUMBER OF SEQ ID NOS: 52
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 3780
;   TYPE: DNA
;   ORGANISM: Zea mays
US-10-209-059-1

  Query Match             35.6%;  Score 1149.8;  DB 12;  Length 3780;
  Best Local Similarity   67.1%;  Pred. No. 0;
  Matches 1735;  Conservative    0;  Mismatches  836;  Indels   15;  Gaps    7;

Qy        374 TGAAATGAATGATGAATATGGGAATCCAATTTGGAAGAATCGGGTGAAGAGCTGTAAGGA 433
Db        797 TGGGCTTAATAGTGTTGACTGGAAGGAAAGAGTTGAGAGCTGGAGGGTTAAACAGGACAA 856

Qy        434 TAAAGAGAACAAGAAGAAAAAGAGAAGTCCTAAGGCTGAAACTGAACCAGCTCAAGTTCC 493
Db        857 AAATATGATGCAAGTGACTAATAAATATCCAGAGGCTAGAGGAGGAGACATGGAGGGGAC 916

Qy        494 TACAGAACAGCAGATGGAAGAGAAACCGTCTGCAGAGGCTTCGGAGCCGCTTTCAATTGT 553
Db        917 TGGCTCAAATGGAGAAGATATGCAAATGGTTGATGATGCACGGCTACCTTTGAGCCGTAT 976

Qy        554 TTATCCAATTCCACGCAACAAGCTCACACCATACAGAGCAGTGATCATTATGCGACTGGT 613
Db        977 CGTGCCAATTTCCTCAAACCAGCTCAACCTTTACCGGGTAGTGATCATTCTCCGTCTTAT 1036

Qy        614 CATTCTGGGCCTCTTCTTCCACTTCAGAATAACAAATCCTGTCGATAGTGCCTTTGGCCT 673
Db       1037 CATCCTGTGCTTCTTCTTCCAGTATCGTGTCAGTCATCCAGTGCGTGATGCTTATGGATT 1096

Qy        674 GTGGCTTACTTCTGTCATATGTGAGATCTGGTTTGCATTTTTTTGGGTGTTGGATCAGTT 733
Db       1097 ATGGCTAGTATCTGTTATCTTGTAGGTCTGGTTTGCCTTTTTGGCTTCTAGATCAGTT 1156

Qy        734 CCCCAAGTGGAATCCTGTCAATATAGAAATTTATATCTAATGCTGTCGGCAAGGTATGA 793
Db       1157 CCCAAAATGGTATCCAATCAACGTGAAATATCTTGATCTGCTTGCATTGAGGTATGA 1216

Qy        794 AAGAGAGGGTGAGCCTTCTCAGCTTGCTGGTGTGGATTTTTCGTGAGTACTGTTGATCC 853
Db       1217 TAGAGAGGGAGAGCCATCACAGCTGGCTCCCATTGATGTCTTCGTCAGTACAGTGGATCC 1276

Qy        854 GCTGAAGGAACCGCCATTGATCACTGCCAATACAGTCCTTTTCATCCTTGCTGTGGACTA 913
Db       1277 ATTGAAGGAACCTCCACTGATCACAGCCAACACTGTTTTGTTCATTCTTTCTGTGGATTA 1336

Qy        914 TCCCGTCGATAAAGTCTCCTGCTACGTGTCTGATGATGGTGCAGCTATGCTTTCATTTGA 973
Db       1337 CCCTGTTGACAAAGTGTCATGCTATGTTTCTGATGATGGTTCAGCTATGCTGACTTTTGA 1396

Qy        974 ATCTCTTGTAGAAACAGCTGAGTTTGCAAGGAAGTGGGTTCCGTTCTGCAAAAAATTCTC 1033
Db       1397 GTCTCTCTCAGAAACCGCAGAATTTGCTAGAAAGTGGGTTCTCTTTTGTAAGAAGCACAA 1456

Qy       1034 AATTGAACCAAGAGCACCGGAGTTTTACTTCTCACAGAAAATTGATTACTTGAAAGACAA 1093
Db       1457 TATTGAACCAAGAGCTCCAGAATTTTACTTTGCTCAAAAAATAGATTACCTGAAGGACAA 1516

Qy       1094 GGTTCAACCTTCTTTCGTGAAAGAACGTAGAGCAATGAAAAAGGATTATGAAGAGTACAA 1153

Db       1517 AATTCAACCTTCATTTGTTAAGGAAAGACGCGCAATGAAGAGGGAGTATGAAGAATTCAA 1576

Qy       1154 AGTCCGAGTTAATGCCCTGGTAGCAAAGGCTCAGAAAACACCTGAAGAAGGATGGACTAT 1213
Db       1577 AGTAAGAATCAATGCCCTTGTTGCCAAAGCACAGAAAGTGCCTGAAGAGGGGTGGACCAT 1636

Qy       1214 GCAAGATGGAACACCTTGGCCTGGGAATAACACACGTGATCACCCTGGGCATGATTCAGG 1273
Db       1637 GGCTGATGGAACTGCATGGCCTGGGAATAATCCTAGGGACCATCCT-GGCATGATTCAGG 1695

Qy       1274 TCTTCCTTGGGAAATACTGGGAGCTCGTGACATTGAAGGAAATGAACTACCTCGTCTAGT 1333
Db       1696 TTTTCTTGGGGCACAGTGGTGGGCTC--GACACTGATGGAAATGAGTTACCACGTCTTGT 1753

Qy       1334 ATATGTCTCCAGGGAGAAGAGACCTGGCTACCAGCACCACAAAAAGGCTGGTGCAGAAAA 1393
Db       1754 CTATGTCTCTCGTGAAAAGAGACCAGGCTTTCAGCATCACAAGAAGGCTGGTGCAATGAA 1813

Qy       1394 TGCTCTGGTGAGAGTGTCTGCAGTACTCACAAATGCTCCCTACATCCTCAATGTTGATTG 1453
Db       1814 TGCGCTGATTCGTGTATCTGCTGTGCTGACAAATGGTGCCTATCTTCTCAATGTGGATTG 1873

Qy       1454 TGATCACTATGTAAACAATAGCAAGGCTGTTCGAGAGGCAATGTGCATCCTGATGGACCC 1513
Db       1874 CGACCATTACTTCAATAGCAGCAAAGCTCTTAGAGAAGCAATGTGCTTCATGATGGATCC 1933

Qy       1514 ACAAGTAGGTCGAGATGTATGCTATGTGCAGTTCCCTCAGAGGTTTGATGGCATAGATAA 1573
Db       1934 GGCTCTAGGAAGGAAAACTTGTTATGTACAATTTCCACAGAGATTTGATGGCATTGACTT 1993

Qy       1574 GAGTGATCGCTACGCCAATCGTAACGTAGTTTTCTTTGATGTTAACATGAAAGGGTTGGA 1633
Db       1994 GCACGATCGATATGCTAATCGGAACATAGTTTTCTTTGATATCAACATGAAAGGTCTGGA 2053

Qy       1634 TGGCATTCAAGGACCAGTATACGTAGGAACTGGTTGTGTTTTCAACAGGCAAGCACTTTA 1693
Db       2054 TGGCATTCAGGGTCCAGTTTACGTGGGAACAGGATGCTGTTTCAATAGACAGGCTTTGTA 2113

Qy       1694 CGGCTACGGGCCTCCTTCTATGCCCAGCTTACGCAAGAGAAAGGATTCTTCATCCTGCTT 1753
Db       2114 TGGATACGATCCTGTTTTGACTGAAGCTGATCTGGAGCCAAACATTGTTATTAAGAGCTG 2173

Qy       1754 CTCATGTTGCTGCCCCTCAAAGAAGAAGCCTGCTCAAGAT-CCAGCTGAGGTATACAGAG 1812
Db       2174 CTGTGGTAGAAGGAAGAAAAAGAACAAGAGTTATATGGATAGTCAAAGCCGTATTATGAA 2233

Qy       1813 ATGCAAAAAGAGAGGATCTCAATGCTGCCATATTTA--ATCTTACAGAGATTGATAATTA 1870
Db       2234 GAGAACAGAATCTTCAGCTCCCATCTTCAATATGGAAGACATCGAAGAGGGTATTGAAGG 2293

Qy       1871 TGACGAGCATGAAAGGTCAATGCTGATCTCCCAGTTGAGCTTTGAGAAAACTTTTGGCTT 1930
Db       2294 TTACGAGGATGAAAGGTCAGTGCTTATGTCCCAGAGGAAATTGGAGAAACGCTTTGGTCA 2353

Qy       1931 ATCTTCTGTCTTCATTGAGTCTACACTAATGGAGAATGGAGGAGTACCCGAGTCTGCCAA 1990
Db       2354 GTCTCCTATTTTCATTGCATCCACCTTTATGACACAAGGTGGCATACCACCTTCAACAAA 2413

Qy       1991 CTCACCACCATTCATCAAGGAAGCGATTCAAGTCATCGGCTGTGGCTATGAAGAGAAGAC 2050
Db       2414 CCCAGCTTCTCTACTAAAGGAAGCTATCCATGTCATCAGTTGTGGATATGAGGACAAAAC 2473

Qy       2051 TGAATGGGGAAAACAGATTGGTTGGATATATGGGTCAGTCACTGAGGATATCTTAAGTGG 2110
Db       2474 TGAATGGGGAAAAGAGATTGGCTGGATCTATGGTTCAGTAACGGAGGATATTCTGACTGG 2533

Qy       2111 CTTCAAGATGCACTGCCGAGGATGGAGATCAATTTACTGCATGCCCGTAAGGCCTGCATT 2170
Db       2534 GTTTAAAATGCATGCAAGGGGCTGGCAATCAATCTACTGCATGCCACCACGACCTTGTTT 2593

Qy       2171 CAAAGGATCTGCACCCATCAACCTGTCTGATAGATTGCACCAGGTCCTCCGATGGGCTCT 2230
Db       2594 CAAGGGTTCTGCACCAATCAATCTTTCCGATCGTCTTAATCAGGTGCTCCGTTGGGCTCT 2653
```

```
2231 TGGTTCTGTGGAAATTTTCTTTAGCAGACACTGTCCCCTCTGGTACGGGTTTGGAGGAGG 2290
2654 TGGGTCAGTGGAAATTCTGCTTAGTAGACATTGTCCTATCTGGTATGGTTACAATGGA-- 2711

2291 CCGTCTTAAATGGCTCCAAAGGCTTGCGTATATAAACACCATTGTGTACCCATTTACATC 2350
2712 -CGATTGAAGCTTTTGGAGAGGCTGGCTTACATCAACACTATTGTATATCCAATCACATC 2770

2351 CCTCCCTCTCATTGCCTATTGCACAATTCCTGCAGTTTGTCTGCTCACCGGAAAATTCAT 2410
2771 CATTCCGCTTATTGCCTATTGTGTGCTTCCCGCTATCTGCCTCCTTACCAATAAATTTAT 2830

2411 CATACCAACGCTCTCAAACCTGGCAAGCATGCTGTTTCTTGGCCTCTTTATCTCCATCAT 2470
2831 CATTCCTGAGATTAGCAATTATGCTGGGATGTTCTTCATTCTTCTTTTCGCCTCCATTTT 2890

2471 TGTAACTGCGGTGCTTGAGCTAAGATGGAGCGGTGTCAGCATTGAAGATTTATGGCGTAA 2530
2891 TGCCACTGGTATATTGGAGCTTAGATGGAGTGGTGTTGGCATTGAAGATTGGTGGAGAAA 2950

2531 TGAACAATTCTGGGTGATCGGAGGTGTTTCAGCCCATCTCTTTGCGGTCTTCCAGGGATT 2590
2951 TGAGCAGTTTTGGGTTATTGGTGGCACCTCTGCCCATCTCTTCGCAGTGTTCCAGGGTCT 3010

2591 CTTAAAAATGTTGGCTGGCATCGATACGAACTTCACTGTCACAGCAAAAGCA---GCCGA 2647
3011 GCTGAAAGTGTTGGCTGGGATTGATACCAACTTCACAGTTACCTCAAAGGCATCTGATGA 3070

2648 AGATGCAGAATTTGGGGAGCTATATATGGTCAAGTGGACAACACTTTTGATTCCTCCAAC 2707
3071 GGATGGCGACTTTGCTGAGCTATATGTGTTCAAGTGGACCAGTTTGCTCATTCCTCCGAC 3130

2708 CACACTTCTCATTATCAATATGTCGGG---TTGTGCTGGATTCTCTGATGCACTCAACAA 2764
3131 CACTGTTCTTGTCATTAACCTGGTCGGAATGGTGGCAGGAATTTCTTATGCCATTAACAG 3190

2765 AGGATATGAAGCATGGGGGCCTCTCTTTGGCAAGGTGTTCTTTGCTTTCTGGGTGATTCT 2824
3191 TGGCTACCAATCCTGGGGTCCGCTCTTTGGAAAGCTGTTCTTCTCGATCTGGGTGATCCT 3250

2825 TCATCTCTATCCATTCCTTAAAGGTCTAATGGGTCGCCAAAACCTAACACCAACCATTGT 2884
3251 CCATCTCTACCCCTTCCTCAAGGGTCTCATGGGAAGGCAGAACCGCACACCAACAATCGT 3310

2885 TGTTCTCTGGTCAGTGCTGTTGGCCTCTGTCTTCTCTCTCGTTTGGGTCAAGATCAATCC 2944
3311 CATTGTCTGGTCCATCCTTCTTGCATCTATCTTCTCCTTGCTGTGGGTGAAGATCGATCC 3370

2945 ATTCGT 2950
3371 TTTCAT 3376
```

```
ESULT 7
S-10-160-719-9
 Sequence 9, Application US/10160719
 Publication No. US20030167528A1
 GENERAL INFORMATION:
  APPLICANT: Kanwarpal S. Dhugga
  APPLICANT:  Timothy G. Helentjaris
  APPLICANT:  Benjamin A.  Bowen
  APPLICANT:  Xun Wang
  TITLE OF INVENTION: Maize Cellulose Synthases and Uses
  TITLE OF INVENTION:  Thereof
  FILE REFERENCE: 0864
  CURRENT APPLICATION NUMBER: US/10/160,719
  CURRENT FILING DATE:  2002-06-03
  PRIOR APPLICATION NUMBER: US/09/371,383A
  PRIOR FILING DATE: 1999-08-06
  PRIOR APPLICATION NUMBER: US 60/096,822
  PRIOR FILING DATE: 1998-08-17
  NUMBER OF SEQ ID NOS: 60
```

```
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 9
;   LENGTH: 3780
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)...(3425)
US-10-160-719-9

  Query Match             45.6%;  Score 2147.8;  DB 12;  Length 3780;
  Best Local Similarity   67.1%;  Pred. No. 0;
  Matches 1735;  Conservative    0;  Mismatches  836;  Indels   15;  Gaps    7;

Qy        374 TGAAATGAATGATGAATATGGGAATCCAATTTGGAAGAATCGGGTGAAGAGCTGTAAGGA 433
Db        797 TGGGCTTAATAGTGTTGACTGGAAGGAAAGAGTTGACAGCTGGAGGGTTAAACAGGACAA 856

Qy        434 TAAAGAGAACAAGAAGAAAAAGCAAGTCGTAAGGCTGAAACTGAACCAGCTCAAGTTCC 493
Db        857 AAATATGATGCAAG-AATAATAAATATCAGAGGCAGAGGAGGAGACATGGAGGGGAC 916

Qy        494 TACAGAACAGCAGA-AAAGAGAAACGTTCAAGCTTCGGAGCCGCTTTCAATTGT 553
Db        917 TGGCTCAAATGGAGAATATGCAAATTTCAATGCAGGCTACTTTGAGCCGTAT 976

Qy        554 TTATCCAATTCCAGCAACAAGCTCACATATACAGAGCAGTGATCATTATGGGACTGGT 613
Db        977 CGTGCCAATTTCCTAAACCAGCTCAACCTTTACCGGGTAGTGATCATTCTCCGTCTTAT 1036

Qy        614 CATTCTGGGCCTCTTCTTCCACTTCAGAATAACAAATCCTGTCGATAGTGCCTTTGGCCT 673
Db       1037 CATCCTGTGCTTCTTCTTCCAGTATCGTGTCAGTCATCCAGTGCGTGATGCTTATGGATT 1096

Qy        674 GTGGCTTACTTCTGTCATATGTGAGACCTGGTTTGCATTTTCTTGGGTGTTGGATCAGTT 733
Db       1097 ATGGCTAGTATCTGTTATCTGCGAGGTCTGGTTTGCCTTGTCTTGGCTTCTAGATCAGTT 1156

Qy        734 CCCCAAGTGGAATCCTGTCAATAGAGAAACGTATATCGAAAGGCTGTCGGCAAGGTATGA 793
Db       1157 CCCAAAATGGTATCTAATCAACCGTGAGACATATCTTGACAGGCTTGCATTGAGGTATGA 1216

Qy        794 AAGAGAGGGTGAGCCTTCTCAGCTTGCTGGTGTCGATTTTTTCGTGAGTACTGTTGATCC 853
Db       1217 TAGAGAGGGAGAGCCATCACAGCTGGCTCCCATTGATGTCTTCGTCAGTACAGTGGATCC 1276

Qy        854 GCTGAAGGAACCGCCATTGATCACTGCCAATACAGTCCTTTCCATCCTTGCTGTGGACTA 913
Db       1277 ATTGAAGGAACCTCCACTGATCACAGCCAACACTGTTTTGTCCATTCTTTCTGTGGATTA 1336

Qy        914 TCCCGTCGATAAAGTCTCCTGCTACGTGTCTGATGATGGTGCAGCTATGCTTTCATTTGA 973
Db       1337 CCCTGTTGACAAAGTGTCATGCTATGTTTCTGATGATGGTTCAGCTATGCTGACTTTTGA 1396

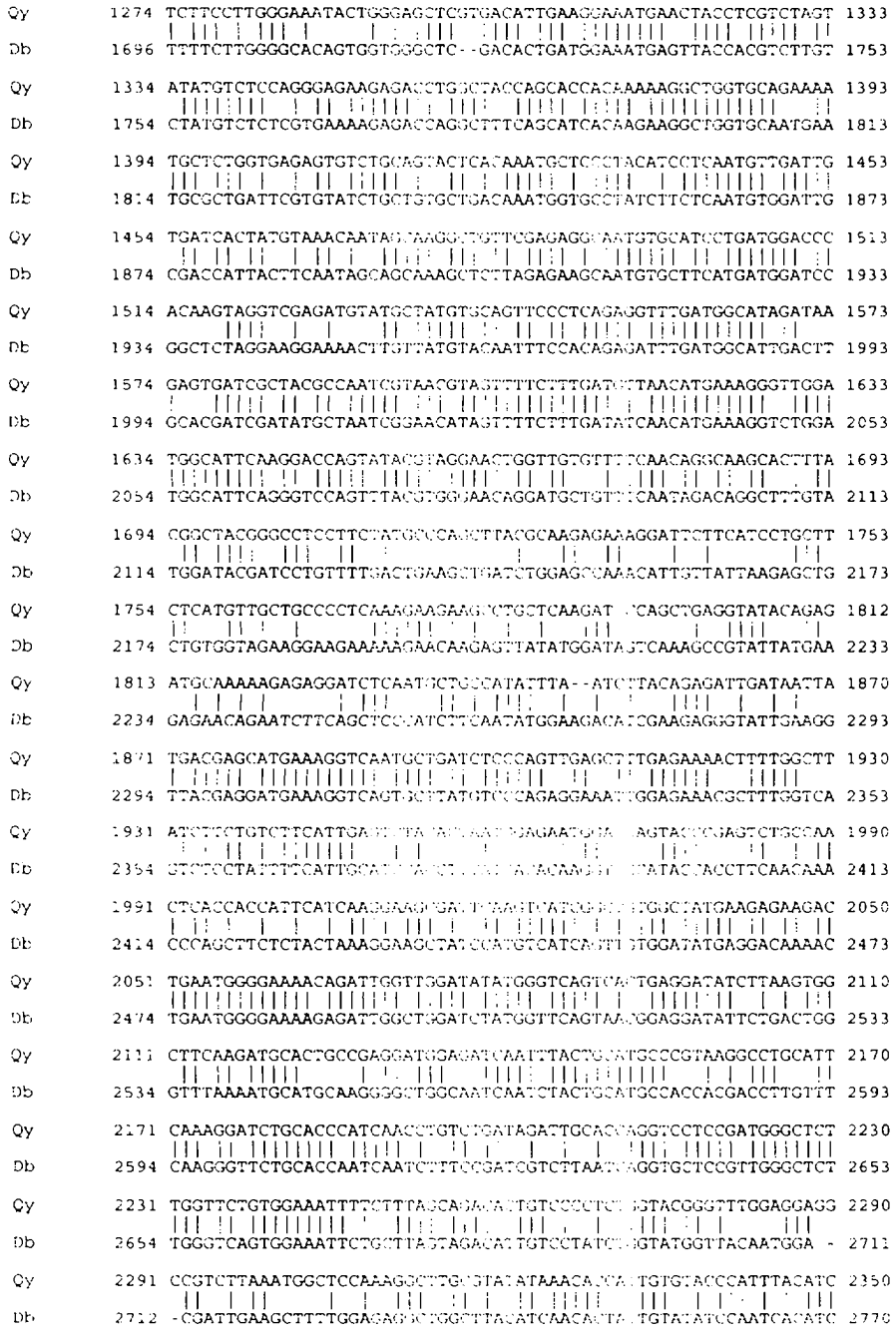
Qy        974 ATCTCTTGTAGAAACAGCTGAGTTTGCAAGGAAGTGGGTTCCGTTCTGCAAAAAATTCTC 1033
Db       1397 GTCTCTCTCAGAAACCGCAGAATTTGCTAGAAAGTGGGTTCCCTTTTGTAAGAAGCACAA 1456

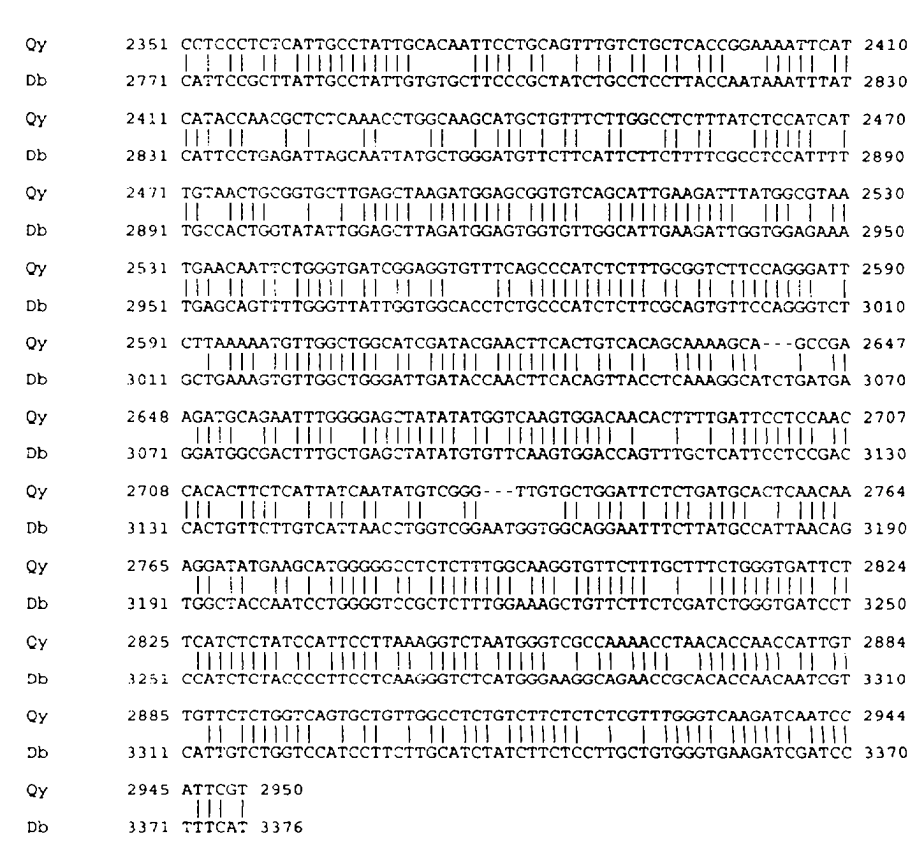
Qy       1034 AATTGAACCAAGAGTACCGGAGTTTTACTTCTCACAGAAAATTGATTACTTGAAAGACAA 1093
Db       1457 TATTGAACCAAGAGTTCCAGAATTTTACTTTGCTCAAAAAATAGATTACCTGAAGGACAA 1516

Qy       1094 GGTTCAACCTTCTTTCGTGAAAGAACGTAGAGCAATGAAAAGGGATTATGAAGAGTACAA 1153
Db       1517 AATTCAACCTTCATTGTTAAGGAAAGACGCGCAATGAAGAGGGAGTATGAAGAATTCAA 1576

Qy       1154 AGTCCGAGTTAATGCCCTGGTAGCAAAGGCTCAGAAAACACCTGAAGAAGGATGGACTAT 1213
Db       1577 AGTAAGAATCAATGCCCTTGTTGCCAAAGCACAGAAAGTGCCTGAAGAGGGGTGGACCAT 1636

Qy       1214 GCAAGATGGAACACCTTGGCCTGGGAATAACACACGTGATCACCCTGGGCATGATTCAGG 1273
Db       1637 GGCTGATGGAACTGCATGGCCTGGGAATAATCCTAGGGACCATCCT-GGCATGATTCAGG 1695
```





RESULT 8
US-10-209-059-13
; Sequence 13, Application US/10209059
; Publication No. US20030163838A1
; GENERAL INFORMATION:
; APPLICANT: Dhugga, Kanwarpal S.
; APPLICANT: Wang, Haiyin
; TITLE OF INVENTION: Maize Cellulose Synthases and Uses
; TITLE OF INVENTION: Thereof
; FILE REFERENCE: 0864R2
; CURRENT APPLICATION NUMBER: US/10/209,059
; CURRENT FILING DATE: 2002-07-31
; PRIOR APPLICATION NUMBER: 60/096,822
; PRIOR FILING DATE: 1998-08-17
; PRIOR APPLICATION NUMBER: 09/371,383
; PRIOR FILING DATE: 1999-08-06
; PRIOR APPLICATION NUMBER: 09/550,483
; PRIOR FILING DATE: 2000-04-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 3704
; TYPE: DNA
; ORGANISM: Zea mays
US-10-209-059-13

Query Match 35.2%; Score 1138.8; DB 12; Length 3704;
Best Local Similarity 67.8%; Pred. No. 2.6e-314;

Matches 1683; Conservative 0; Mismatches 783; Indels 15; Gaps 6;

```
      518 ACCGTCTGCAGAGGCTTCGGAGCCGCTTTCAATTGTTTATCCAATTCCACGCAACAAGCT 577
          || |  ||  ||  ||  | | || || || |       |||||||      ||   |
     1018 ACTGAATGATGAAACTCGGCAACCTCTATCTAGAAAAGTGCCAATTCCTTCATCCAGAAT 1077

      578 CACACCATACAGAGCAGTGATCATTATGCGACTGGTCATTCTGGGCCCTCTTCTTCCACTT 637
          | || ||||||   || ||  |  | ||  |||   ||||  || | ||||| | ||
     1078 AAATCCGTACAGAATGGTCATTGTGCTACGTTTGGCTGTTCTATGCATATTCTTGCGCTA 1137

      638 CAGAATAACAAATCCTGTCGATAGTGCCTTTGGCCTGTGGCTTACTTCTGTCATATGTGA 697
          | | || ||| |||||||  | | ||| | |  ||||||||||   || ||||||||||
     1138 CCGTATCACACATCCTGTGAACAATGCATATCCACTGTGGCTTTTATCCGTCATATGTGA 1197

      698 GATCTGGTTTGCATTTTCTTGGGTGTTGGATCAGTTCCCCAAGTGGAATCCTGTCAATAG 757
          ||||||||||||| || || ||| | ||||||||||||||| ||||||  || ||||  |
     1198 GATCTGGTTTGCTTTGTCCTGGATTTTGGATCAGTTCCCAAAGTGGTCCCCAATCAACCG 1257

      758 AGAAACGTATATCGAAAGGCTGTCGGCAAGGTATGAAAGAGAGGGTGAGCCTTCTCAGCT 817
          |||||  ||  | || || ||| |  |||||||||  ||| |||||  || ||||  |
     1258 TGAAACATACCTTGATAGACTGGCTTTAAGGTATGACCGAGAAGGTGAACCATCTCAATT 1317

      818 TGCTGGTGTGGATTTTTTCGTGAGTACTGTTGATCCGCTGAAGGAACCGCCATTGATCAC 877
          |||  ||| ||| |||| || ||||||||  ||||  |||||||  || ||   |  |
     1318 AGCTCCTGTTGATATTTTTGTCAGTACTGTGGATCCAATGAAGGAGCCTCCTCTTGTCAC 1377

      878 TGCCAATACAGTCCTTTCCATCCTTGCTGTGGACTATCCCGTCGATAAAGTCTCCTGCTA 937
          ||| ||||| || ||||||||||||||||| || ||||| || || || || || |||||
     1378 TGCAAATACTGTGCTTTCCATCCTTGCTGTCGATTATCCGGTTGACAAGGTATCTTGCTA 1437

      938 CGTGTCTGATGATGGTGCAGCTATGCTTTCATTTGAATCTCTTGTAGAAACAGCTGAGTT 997
          || || |||||||| || ||||||  | |||||  ||||  |||||  | |||||
     1438 TGTTTCGGATGATGGAGCTGCTATGCTGACTTTTGATGCTCTCTCTGAAACTTCAGAGTT 1497

      998 TGCAAGGAAGTGGGTTCCGTTCTGCAAAAAATTCTCAATTGAACCAAGAGCACCGGAGTT 1057
          ||| || || |||||||||||||| || || | |  || || || || || ||||| |
     1498 TGCTAGAAAATGGGTTCCGTTCTGTAAGAAGTACAACATAGAGCCTAGGGCCCCGGAATG 1557

     1058 TTACTTCTCACAGAAAATTGATTACTTGAAAGACAAGGTTCAACCTTCTTTCGTGAAAGA 1117
          |||||  | |||||||||||||||||||||||||||||| |||||| | || || ||||||||
     1558 GTACTTTGCTCAGAAAATTGATTACTTGAAAGACAAAGTTCAAACCTCATTTGTGAAAGA 1617

     1118 ACGTAGAGCAATGAAAAGGGATTATGAAGAGTACAAAGTCCGAGTTAATGCCCTGGTAGC 1177
          |||  | || |||||  || ||  |||||||  | ||||||  ||  | ||||  || |||||
     1618 ACGCCGGGCCATGAAGAGAGAATATGAAGAATTCAAAGTTCGTATCAATGGTCTTGTAGC 1677

     1178 AAAGGCTCAGAAAACACCTGAAGAAGGATGGACTATGCAAGATGGAACACCTTGGCCTGG 1237
          |||||  || |||   || || || |||||||  ||||||||||| ||||||||||||||
     1678 CAAGGCACAAAAAGTTCCCGAGGAGGGATGGATCATGCAAGATGGTACACCTTGGCCTGG 1737

     1238 GAATAACACACGTGATCACCCTGGGCATGATTCAGGTCTTCCTTGGGAAATACTGGGAGC 1297
          ||| || ||  | || || ||| || ||||||||||| |||| ||  |     ||||
     1738 GAACAATACTAGGGACCATCCT-GGAATGATTCAGGTTTTCCTGGGTCA--CAGTGGAGG 1794

     1298 TCGTGACATTGAAGGAAATGAACTACCTCGTCTAGTATATGTCTCCAGGGAGAAGAGACC 1357
          | |||| |||||||  |||||| | || |||| ||  | || |||||||  ||||||
     1795 GCTTGACGTTGAAGGCAATGAACTTCCTCGTTTGGTTTATGTGTCTCGTGAAAAACGTCC 1854

     1358 TGGCTACCAGCACCACAAAAAGGCTGGTGCAGAAAATGCTCTGGTGAGAGTGTCTGCAGT 1417
          ||| | ||| || |||||  ||||||||||     |||||  || ||  | || || || ||
     1855 TGGATTCCAACATCACAAGAAGGCTGGTGCCATGAATGCACTTGTTCGTGTATCAGCTGT 1914

     1418 ACTCACAAATGCTCCCTACATCCTCAATGTTGATTGTGATCACTATGTAAACAATAGCAA 1477
          || || ||||  |  |||||  | ||| ||||||||||| |||||  | || ||||||||
     1915 CCTTACTAATGGGCAATACATGTTGAATCTTGATTGTGACCACTACATCAATAATAGCAA 1974

     1478 GGCTGTTCGAGAGGCAATGTGCATCCTGATGGACCCACAAGTAGGTCGAGATGTATGCTA 1537
          |||| |||||||  || ||||||  ||| |||||||||  |  |||| |  |||| || ||
     1975 GGCTCTTCGAGAAGCTATGTGCTTCCTTATGGACCCAAACCTAGGAAGGAATGTCTGTTA 2034

     1538 TGTGCAGTTCCCTCAGAGGTTTGATGGCATAGATAAGAGTGATCGCTACGCCAATCGTAA 1597
          ||| || || |||||||||||||||||| || |||| || ||| || || || ||  | ||
     2035 TGTCCAATTTCCTCAGAGGTTTGATGGTATTGATAGGAATGACCGATATGCAAACAGGAA 2094
```

```
Qy    1598 CGTAGTTTTCTTTGACTTTAACATAAAGTGTTGGATGGCATTCAAGGACCAGTATACGT 1657
           |  || |  |  ||| || |  | | | ||||||||| || | || |
Db    2095 CACTGTGTTTTTGGACTTAACTTAAATGTCTTGACGGCATTCAAGGGCCAGTTTATGT 2154

Qy    1658 AGGAACTGGTTGTGTTCTCAACAGAAAGACTTTAGGCTACGGGCCTCCTTCTATGCC 1717
           ||||||||||||| | ||||  |  | || || | |||||     | |
Db    2155 GGGAACTGGTTGTGTTTTAACAGAAAGACTTTATATGGTTATGAGCCTCCAGTCAAGAA 2214

Qy    1718 CAGCTTACGCAAGAGAAAGGATTC  TTCATCCTGCTTCTCATGTTGCTGCCCCTCAAA 1774
           |  |  |||  ||  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db    2215 AAAAAAGCCAGGCT TTCTCTTCGTTTTTGGGGGAAGGAAAAAGACGTCAAAATCTAA 2274

Qy    1775 GAAGAAGCCTGCTCAAGATCCAGCTGAGTATACAGAGATGCAAAAAGAGAGGATCTCAA 1834
           |||||  | |  |  |  |  ||  |  |  |  |  |  |  |  |  | |  |
Db    2275 GAAGAGCTCGGAAAAGAAGAAGTCAGATAGACATGCAGACAGTTCTGTACCAGTATTTAA 2334

Qy    1835 TGCTGCTATATTTAAGTTTACAGAGATCTGTATAATTATGACGAGCATGAAAGGTCAATGCT 1894
           |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db    2335 TCTCGAAGATATAGTAGGAGATAAGTTCTTTAGTTTTGATGATGAGAAATTGCTGAT 2394

Qy    1895 GATCTCCTAGTTGATAAGAAAGTCAAGTTCTGTCTTCATTTAGTCTAC 1954
           || | |  ||  |  |||  |  |  |  |  |  |||  |  |||
Db    2395 TATGTCTCAAATGAGGAGAAAGAAGTCAAGTGTTTTTGTAAGTCTAC 2454

Qy    1955 ACTAATGGAGAATGCAGGAGTACCCAAGTCTGCAACTCACCACCATTCATCAAGGAAGC 2014
           || |||||  |||  | |  ||  |||  |  |  |  | | | |  |||| |
Db    2455 TCTGATGGAATATGTTGGTGTTCCACAATCTGTAACTCCAGAGTCTCTTCTCAAAGAAGC 2514

Qy    2015 GATTCAAGTCATCGTCTGTGGCTATGAAGAGAAGATTGAATGGGGAAAACAGATTGGTTG 2074
           |||||  |||  |  |||||  || |  ||  |||  |||||  |||  |  || |
Db    2515 TATTCATGTCATCAGTTGTGGCTATGAGGACAAAAGTGACTGGGGAACTGAGATTGGGTG 2574

Qy    2075 GATATATGGGTCAGTCACTGAGGATATGTTAAATGTTTTCAAGATGCACTGCCGAGGATG 2134
           ||| |||  || |  || | |  |  |  |  |  |||||||||  |||  | ||
Db    2575 GATCTATGGTTCTGTGACAGAAGACATTCTCATGGATTCAAGATGCATGCTCGAGGCTG 2634

Qy    2135 GAGATCAATTTACTGCATGCCGTAAGCCCTGGATTCAAAGGATCTGCACCCATCAACCT 2194
           |||:||| :|||  |||||         || |  ||| | |||||||  | |||||||
Db    2635 GCGATCAATCTACTGGATGCCTAAGTAACCAGTTTCAAGGGATCTGCTCCTATCAACCT 2694

Qy    2195 GTCTGATAGATTGCACCAGGTCCTCCGATGGGCTCTTGGTTCTGTGGAAATTTTCTTTAG 2254
           ||  ||  | ||  | || | ||  |  |||  ||||.|||  | |||| | |  |||
Db    2695 TTCGGATCGTTTGAATCAAGTGCTTCGGTGGGCTCTTGGTTCCATTGAAATTCTTTTCAG 2754

Qy    2255 CAGACACTGTCCCCTCTGGTACGGGTTTGGAGAAGGCGTCTTAAATGGCTCCAAAGGCT 2314
           |||  || |||||  | |:|||  |  |  |||  ||| ||||||  ||  | ||  |
Db    2755 CAGGCATTGTCCCATATGGTATGGCTAT- -GGAGGCGGCTTAAATTCCTGGAGAGATT 2811

Qy    2315 TGCGTATATAAACACCATTGTGTACCCATTTACATCCCTCCCTCTCATTGCCTATTGCAC 2374
           |||  |||||  |||:  |  |  ||  :|||'  ||||  |||  |    || ||||
Db    2812 TGCTTATATCAACACAACAATTTATCCACTCA'ATCAATCCCGCTCCTCCTGTACTGCAT 2871

Qy    2375 AATTCCTGCAGTTTGTCTGCTCACCGGAAAATTCATCATACCAACGCTCTCAAACCTGGC 2434
           | | ||  |||||||.||  |||  |  || |||| | |  |||||  | |
Db    2872 ATTGCCAGCAGTTTGTCTTCTCACTCGGAAGTTCATCATCCCAAAGATTAGTAACCTAGA 2931

Qy    2435 AAGCATGGTGTTTCTTGGCCTCTTTATGTCCATCATTGTAACTGCGGTGCTTGAGCTAAG 2494
           ||  |  |||| |  |||||||:|| |.|| ||  ||||   | |:|||  : ||
Db    2932 GAGTGTTTGGTTTATATCGCTCTTTATGTCAATGTTTGCCACTGGTATCCTTGAGATGAG 2991

Qy    2495 ATGGAGCGGTGTCATGATTGAAGATTTATGGCGTAATGAACAATTCTGGGTGATCGGAGG 2554
           |||||  |||||  |  |||| |.  ||, ' |,  | |, ||||:||| || || ||
Db    2992 GTGGAGTGGTGTTGTGCATTGATGAATGGTGGAGGAACGAGCAGTTCTGGGTCATTGGTGG 3051

Qy    2555 TGTTTCAGCCCATCTCTTTGCGGTCTTCCAGGGATTCTTAAAAATGTTGGCTGGCATCGA 2614
           | |||  || |||  | |||  |||| |||  || | || |  || | ||||  |||
Db    3052 TATTTCTGCGCATTTATTTGCCGTCTTCCAGGGTCTCCTGAAGGTGCTTGCTGGTATCGA 3111

Qy    2615 TACGAACTTCACTGTCACAGCAAAAG---CAGCCGAAGATGCAGAATTTGGGGAGCTATA 2671
           |||| |||||||||||  | || |  | |||||| |  || |||  |||| ||
Db    3112 CACGAGCTTCACTGTCACCTCTAAGGCCACTGACGAAGAAGGTGATTTTGCCGAGCTCTA 3171
```

```
Qy       2672 TATGGTCAAGTGGACAACACTTTTGATTCCTCCAACCACACTTCTCATTATCAATATGTC 2731
Db       3172 CATGTTCAAGTGGACAACGCTTCTGATCCCACCAACCACTATTTTGATCATCAACCTGGT 3231

Qy       2732 GGG---TTGTGCTGGATTCTCTGATGTACTCAACAAAGGATATGAAGCATGGGGGCCTCT 2788
Db       3232 CGGCGTGGTCGCTGGCATTTCCTACGCAATCAATAGCGGTTACCAGTCATGGGGACCTCT 3291

Qy       2789 CTTTGGCAAGGTGTTCTTTGCTTTCTGGGTGATTCTTCATCTCTATCCATTCCTTAAAGG 2848
Db       3292 TTTCGGGAAGCTCTTCTTTGGGTTCTGGGTGATTGTCCACTTGTACCCCTTCCTCAAGGG 3351

Qy       2849 TCTAATGGGTCGCCAAAACCTAACACCAACCATTGTTGTTCTCTGGTCAGTGCTGTTGGC 2908
Db       3352 CCTCATGGGGAAGCAGAACCGCAGGCGGACCATTGTCGTTGTCTGGGCTATCCTCCTTGC 3411

Qy       2909 CTCTGTCTTCTCTCTCGTTTGGGTCAAGATCAATCCATTCGTTAACAAAGTTGATAACAC 2968
Db       3412 GTCGATCTTTTCCCTGATGTGGGTTCGTATCGATCCATTCAACACCCGGGTCACTGGCCC 3471

Qy       2969 CTTGGTTGCGGAGACCTGCAT 2989
Db       3472 TGATATCGCGAAATGTGGCAT 3492


RESULT 9
US-10-160-719-57
; Sequence 57, Application US/10160719
; Publication No. US20030167528A1
; GENERAL INFORMATION:
;  APPLICANT: Kanwarpal S. Dhugga
;  APPLICANT:  Timothy G. Helentjaris
;  APPLICANT:  Benjamin A.  Bowen
;  APPLICANT:  Xun Wang
;  TITLE OF INVENTION: Maize Cellulose Synthases and Uses
;  TITLE OF INVENTION:  Thereof
;  FILE REFERENCE: 0864
;  CURRENT APPLICATION NUMBER: US/10/160,719
;  CURRENT FILING DATE:  2002-06-03
;  PRIOR APPLICATION NUMBER: US/09/371,383A
;  PRIOR FILING DATE: 1999-08-06
;  PRIOR APPLICATION NUMBER: US 60/096,822
;  PRIOR FILING DATE: 1998-08-17
;  NUMBER OF SEQ ID NOS: 60
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 57
;  LENGTH: 3704
;  TYPE: DNA
;  ORGANISM: Zea mays
;  FEATURE:
;  NAME/KEY: CDS
;  LOCATION: (272)...(3499)
US 10-160-719-57

  Query Match             35.2%;  Score 1139.8;  DB 10;  Length 3704;
  Best Local Similarity   67.8%;  Pred. No. 2.6e-314;
  Matches 1683;  Conservative    0;  Mismatches  783;  Indels   15;  Gaps    6;

Qy        518 ACCGTCTGCAGAGGCTTCGGAGCGGCTTCAATTGTTTATCTAATTCCACGCAACAAGCT 577
Db       1018 ACTGAATGATGAAACTCGGCAACCTCTATCTAGAAAAGTGGAATTCCTTCATCCAGAAT 1077

Qy        578 CACACCATACAGAGCAGTGATCATTATGCGACTGGTCATTCTGGGCCTCTTCTTCCACTT 637
Db       1078 AAATCCGTACAGAATGGTCATTGTGCTACGTTTGGCTGTTCTATGCATATTCTTGCGCTA 1137

Qy        638 CAGAATAACAAATCCTGTCGATAGTGCTTTGGCCTGTGGCTTACTTCTGTCATATGTGA 697
Db       1138 CCGTATCACACATCCTGTGAACAATGCATATCCACTGTGGCTTTATCCGTCATATGTGA 1197

Qy        698 GATCTGGTTTGCATTTTCTTGGGTGTTGATCAGTTCCCCAAGTGGAATCCTGTCAATAG 757
Db       1198 GATCTGGTTTGCTTTGTCCTGGACTTTGGATCAGTTCCCAAAGTGGTCCCCAATCAACTG 1257

Qy        758 AGAAACGTATATCGAAAGGCTGTCGGCAAGGTATGAAAGAGAGGGTGAGCCTTCTCAGCT 817
Db       1258 TGAAACATACCTTGATAGACTGGCTTTAAGGTATGACCGAGAAGGTGAACCATCTCAATT 1317

Qy        818 TGCTGGTGTGGATTTTTTCGTGAGTACTGTTGATCCGCTGAAGGAACCGCCATTGATCAC 877
Db       1318 AGCTCCTGTTGATATTTTTGTCAGTACTGTGGATCCAATGAAGGAGCCTCCTCTTGTCAC 1377

Qy        878 TGCCAATACAGTCCTTTCCATCCTTGCTGTGGACTATCCCGTCGATAAAGTCTCCTGCTA 937
Db       1378 TGCAAATACTGTGCTTTCCATCCTTGCTGTCGATTATCCGGTTGACAAGGTATCTTGCTA 1437

Qy        938 CGTGTCTGATGATGGTGCAGCTATGCTTTCATTTGAATCTCTTGTAGAAACAGCTGAGTT 997
Db       1438 TGTTTCGGATGATGGAGCTGCTATGCTGACTTTTGATGCTCTCTCTGAAACTTCAGAGTT 1497

Qy        998 TGCAAGGAAGTGGGTTCCGTTCTGCAAAAAATTCTCAATTGAACCAAGAGCACCGGAGTT 1057
Db       1498 TGCTAGAAAATGGGTTCCGTTCTGTAAGAAGTACAACATAGAGCCTAGGGCCCCGGAATG 1557

Qy       1058 TTACTTCTCACAGAAAATTGATTACTTGAAAGACAAGGTTCAACCTTCTTTCGTGAAAGA 1117
Db       1558 GTACTTTGCTCAGAAAATTGATTACTTGAAAGACAAAGTTCAAACCTCATTTGTGAAAGA 1617

Qy       1118 ACGTAGAGCAATGAAAAGGGATTATGAAGAGTACAAAGTCCGAGTTAATGCCCTGGTAGC 1177
Db       1618 ACGCCGGGCCATGAAGAGAGAATATGAAGAATTCAAAGTTCGTATCAATGGTCTTGTAGC 1677

Qy       1178 AAAGGCTCAGAAAACACCTGAAGAAGGATGGACTATGCAAGATGGAACACCTTGGCCTGG 1237
Db       1678 CAAGGCACAAAAAGTTCCCGAGGAGGGATGGATCATGCAAGATGGTACACCTTGGCCTGG 1737

Qy       1238 GAATAACACACGTGATCACCCTGGGCATGATTCAGGTCTTCCTTGGGAAATACTGGGAGC 1297
Db       1738 GAACAATACTAGGGACCATCCT-GGAATGATTCAGGTTTTCCTGGGTCA--CAGTGGAGG 1794

Qy       1298 TCGTGACATTGAAGGAAATGAACTACCTCGTCTAGTATATGTCTCCAGGGAGAAGAGACC 1357
Db       1795 GCTTGACGTTGAAGGCAATGAACTTCCTCGTTTGGTTTATGTGTCTCGTGAAAAACGTCC 1854

Qy       1358 TGGCTACCAGCACCACAAAAAGGCTGGTGCAGAAAATGCTCTGGTGAGAGTGTCTGCAGT 1417
Db       1855 TGGATTCCAACATCACAAGAAGGCTGGTGCCATGAATGCACTTGTTCGTGTATCAGCTGT 1914

Qy       1419 ACTCACAAATGCTCCCTACATCCTCAATGTTGATTGTGATCACTATGTAAACAATAGCAA 1477
Db       1915 CCTTACTAATGGGCAATACATGTTGAATCTTGATTGTGACCACTACATCAATAATAGCAA 1974

Qy       1478 GGCTGTTCGAGAGGCAATGTGCATCCTGATGGACCCACAAGTAGGTCGAGATGTATGCTA 1537
Db       1975 GGCTCTTCGAGAAGCTATGTGCTTCCTTATGGACCCAAACCTAGGAAGGAATGTCTGTTA 2034

Qy       1538 TGTGCAGTTCCCTCAGAGGTTTGATGGCATAGATAAGAGTGATCGCTACGCCAATCGTAA 1597
Db       2035 TGTCCAATTTCCTCAGAGGTTTGATGGTATTGATAGGAATGACCGATATGCAAACAGGAA 2094

Qy       1598 CGTAGTTTTCTTTGATGTTAACATGAAAGGGTTGGATGGCATTCAAGGACCAGTATACGT 1657
Db       2095 CACTGTGTTTTTCGATATTAACTTGAGAGGTCTTGACGGCATTCAAGGGCCAGTTTATGT 2154

Qy       1658 AGGAACTGGTTGTGTTTTCAACAGGCAAGCACTTTACGGCTACGGGCCTCCTTCTATGCC 1717
Db       2155 GGGAACTGGTTGTGTGTTTAACAGAACGGCCTTATATGGTTATGAGCCTCCAGTCAAGAA 2214

Qy       1718 CAGCTTACGCAAGAGAAAGGATTC---TTCATCCTGCTTCTCATGTTGCTGCCCCTCAAA 1774
Db       2215 AAAAAAGCCAGGCTTCTTCTCTTCGCTTTGTGGGGGAAGGAAAAAGACGTCAAAATCTAA 2274

Qy       1775 GAAGAAGCCTGCTCAAGATCCAGCTGAGGTATACAGAGATGCAAAAAGAGAGGATCTCAA 1834
Db       2275 GAAGAGCTCGGAAAAGAAGAAGTCACATAGACACGCAGACAGTTCTGTACCAGTATTTAA 2334
```

```
1835 TGCTGCCATATTTAATCTTACAGAGATTGATAATTATGACGAGCATGAAAGGTCAATGCT 1894
2335 TCTCGAAGATATAGAGGAAGGGATTGAAGGTTCTCAGTTTGATGATGAGAAATCGCTGAT 2394

1895 GATCTCCCAGTTGAGCTTTGAGAAAACTTTTGGCTTATCTTCTGTCTTCATTGAGTCTAC 1954
2395 TATGTCTCAAATGAGCTTGGAGAAGAGATTTGGCCAGTCCAGTGTTTTTGTAGCCTCTAC 2454

1955 ACTAATGGAGAATGGAGGAGTACCCGAGTCTGCCAACTCACCACCATTCATCAAGGAAGC 2014
2455 TCTGATGGAATATGGTGGTGTTCCACAATCTGCAACTCCAGAGTCTCTTCTGAAAGAAGC 2514

2015 GATTCAAGTCATCGGCTGTGGCTATGAAGAGAAGACTGAATGGGGAAAACAGATTGGTTG 2074
2515 TATTCATGTCATCAGCTGTGGCTATGAGGACAAAACTGACTGGGGAACTGAGATTGGGTG 2574

2075 GATATATGGGTCAGTCACTGAGGATATCTTAAGTGGCTTCAAGATGCACTGCCGAGGATG 2134
2575 GATCTATGGTTCTGTTACAGAAGACATTCTCACCGGATTCAAGATGCATGCTCGAGGCTG 2634

2135 GAGATCAATTTACTGCATGCCCGTAAGGCCTGCATTCAAAGGATCTGCACCCATCAACCT 2194
2635 GCGATCAATCTACTGCATGCCTAAGCGACCAGCTTTCAAGGGATCTGCTCCTATCAACCT 2694

2195 GTCTGATAGATTGCACCAGGTCCTCCGATGGGCTCTTGGTTCTGTGGAAATTTTCTTTAG 2254
2695 TTCGGATCGTTTGAATCAAGTGCTTCGGTGGGCTCTTGGTTCCATTGAAATTCTTTTCAG 2754

2255 CAGACACTGTCCCCTCTGGTACGGGTTTGGAGGAGGCCGTCTTAAATGGCTCCAAAGGCT 2314
2755 CAGGCATTGTCCCATATGGTATGGCTAT---GGAGGCCGGCTTAAATTCCTGGAGAGATT 2811

2315 TGCGTATATAAACACCATTGTGTACCCATTTACATCCCTCCCTCTCATTGCCTATTGCAC 2374
2812 TGCTTATATCAACACAACAATTTATCCACTCACATCAATCCCGCTCCTCCTGTACTGCAT 2871

2375 AATTCCTGCAGTTTGTCTGCTCACCGGAAAATTCATCATACCAACGCTCTCAAACCTGGC 2434
2872 ATTGCCAGCAGTTTGTCTTCTCACTGGGAAGTTCATCATCCCAAAGATTAGTAACCTAGA 2931

2435 AAGCATGCTGTTTCTTGGCCTCTTTATCTCCATCATTGTAACTGCGGTGCTTGAGCTAAG 2494
2932 GAGTGTTTGGTTTATATCGCTCTTTATCTCAATCTTTGCCACTGGTATCCTTGAGATGAG 2991

2495 ATGGAGCGGTGTCAGCATTGAAGATTTATGGCGTAATGAACAATTCTGGGTGATCGGAGG 2554
2992 GTGGAGTGGTGTTGGCATTGATGAATGGTGGAGGAACGAGCAGTTCTGGGTCATTGGTGG 3051

2555 TGTTTCAGCCCATCTCTTTGCGGTCTTCCAGGGATTCTTAAAAATGTTGGCTGGCATCGA 2614
3052 TATTTCTGCGCATTTATTTGCCGTCTTCCAGGGTCTCCTGAAGGTGCTTGCTGGTATCGA 3111

2615 TACGAACTTCACTGTCACAGCAAAAG---CAGCCGAAGATGCAGAATTTGGGGAGCTATA 2671
3112 CACGAGCTTCACTGTCACCTCTAAGGCCACTGACGAAGAAGGTGATTTTGCCGAGCTCTA 3171

2672 TATGGTCAAGTGGACAACACTTTTGATTCCTCCAACCACACTTCTCATTATCAATATGTC 2731
3172 CATGTTCAAGTGGACAACGCTTCTGATCCCACCAACCACTATTTTGATCATCAACCTGGT 3231

2732 GGG---TTGTGCTGGATTCTCTGATGCACTCAACAAAGGATATGAAGCATGGGGGCCTCT 2788
3232 CGGCGTGGTCGCTGGCATTTCCTACGCAATCAATAGCGGTTACCAGTCATGGGGACCTCT 3291

2789 CTTTGGCAAGGTGTTCTTTGCTTTCTGGGTGATTCTTCATCTCTATCCATTCCTTAAAGG 2848
3292 TTTCGGGAAGCTCTTCTTTGCGTTCTGGGTGATTGTCCACCTGTACCCCTTCCTCAAGGG 3351

2849 TCTAATGGGTCGCCAAAACCTAACACCAACCATTGTTGTTCTCTGGTCAGTGCTGTTGGC 2908
3352 CCTCATGGGGAAGCAGAACCGCACGCCGACCATTGTCGTTGTCTGGGCTATCCTCCTTGC 3411

2909 CTCTGTCTTCTCTCTCGTTTGGGTCAAGATCAATCCATTCGTTAACAAAGTTGATAACAC 2968
```

```
Db       3412 GTCGATCTTTTCCCTGATGTGGGTTCGTATCGATCCATTCACCACCCGGGTCACTGGCCC 3471

Qy       2969 CTTGGTTGCGGAGACCTGCAT 2989
Db       3472 TGATATCGCGAAATGTGGCAT 3492


RESULT 10
US-10-209-059-45
; Sequence 45, Application US/10209059
; Publication No. US20030163838A1
; GENERAL INFORMATION:
;  APPLICANT: Dhugga, Kanwarpal S.
;  APPLICANT:  Wang, Haiyin
;  TITLE OF INVENTION: Maize Cellulose Synthases and Uses
;  TITLE OF INVENTION:  Thereof
;  FILE REFERENCE: 0864R2
;  CURRENT APPLICATION NUMBER: US/10/209,059
;  CURRENT FILING DATE: 2002-07-31
;  PRIOR APPLICATION NUMBER: 60/308,822
;  PRIOR FILING DATE: 1998-08-03
;  PRIOR APPLICATION NUMBER: 09/371,580
;  PRIOR FILING DATE: 1999-08-06
;  PRIOR APPLICATION NUMBER: 09/550,485
;  PRIOR FILING DATE: 2000-04-14
;  NUMBER OF SEQ ID NOS: 52
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 45
;   LENGTH: 3725
;   TYPE: DNA
;   ORGANISM: Zea mays
US-10-209-059-45

  Query Match           35.2%;  Score 1137.6;  DB 12;  Length 3725;
  Best Local Similarity  69.0%;  Pred. No. 5.2e-314;
  Matches 1662;  Conservative    0;  Mismatches  729;  Indels   17;  Gaps    7;

Qy        553 TTTATCCAATTCCATGCAACAAGCTCACACCATACAGAGCAGTGATCATTATGCGACTGG 612
Db        951 TTGTGCCAATTTCCTCAAACCAGCTCAACGTTTACCGGATAGTAATCATTCTCCGTCTTA 1010

Qy        613 TCATTCTGGGCCTCTTCTTCCACTTCAGAATAACAAATCCTGTCGATAGTGCCTTTGGCC 672
Db       1011 TCATCCTGTGCTTCTCTTCCAATATCGTATCAGTCATCCAGTGCGTAATGCTTATGGAT 1070

Qy        673 TGTGGCTTACTTCTGTCATATGTGAGATCTGGTTTGCATTTTCTTGGGTGTTGGATCAGT 732
Db       1071 TGTGGCTAGTATCTGTTATCTGTGAGGTCTGGCTTGCCTTGTCCTGGCTTCTAGATCAGT 1130

Qy        733 TCCCCAAGTGGAATCCTGTCAATAGAGAAACGTATATCGAAAGGCTGTCGGCAAGGTATG 792
Db       1131 TCCCAAAATGGTATCCAATCAACCGTGAGACATATGTCGACAGGCTTGCATTGAGGTATG 1190

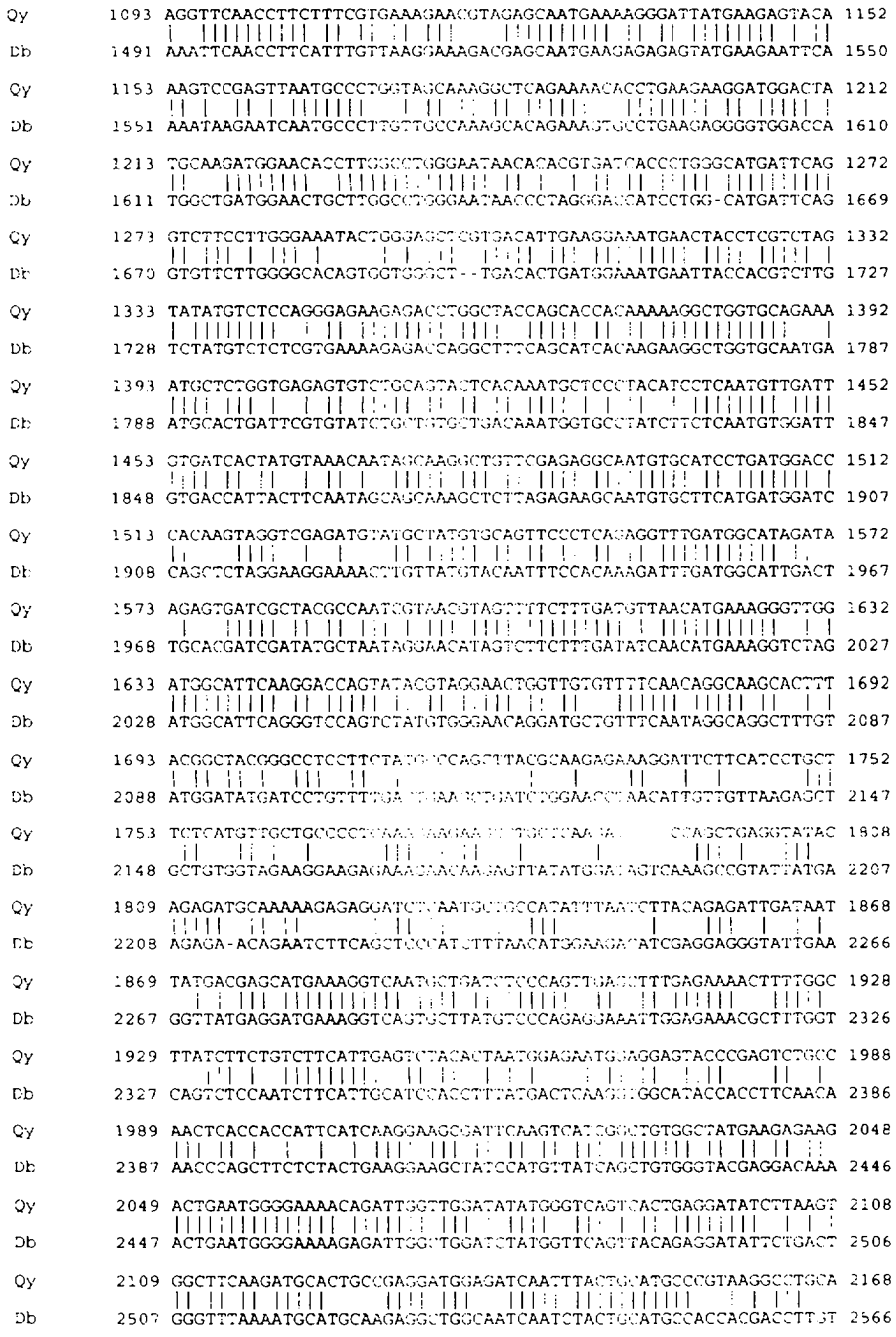
Qy        793 AAAGAGAGGGTGAGCTCTTCTCAGCTTGCTGGTTGGATTTTTTCGTGAGTACTGTTGATC 852
Db       1191 ATAGAGAGGGAGAGTCATCACAGCTGGCTCCATTGATGTCTTTGTCAGTACAGTGGATC 1250

Qy        853 CGCTGAAGGAACCGTCATTGATCACTGCCAATACAGTCCTTTCCATCCTTGCTGTGGACT 912
Db       1251 CATTGAAGGAACCTTCACTGATCACAGTGAACACTGTTTTGTCCATTCTTGCTGTGGATT 1310

Qy        913 ATCCCGTCGATAAAGTCTCCTGCTACGTGTCTGATGATGGTGCAGCTATGCTTTCATTTG 972
Db       1311 ACCCTGTTGACAAAGTGTCATGCTATGTTTCTTGATGATGGCTCAGCTATGCTGACTTTTG 1370

Qy        973 AATCTCTTGTAGAAACAGCTGAGTTTGCAAGGAAGTGGGTTCCGTTCTGCAAAAAATTCT 1032
Db       1371 AGTCTCTCTCTGAAACTGCCGAATTTGCTAGAAAGTGGGTTCCCTTTTGTAAGAAGCACA 1430

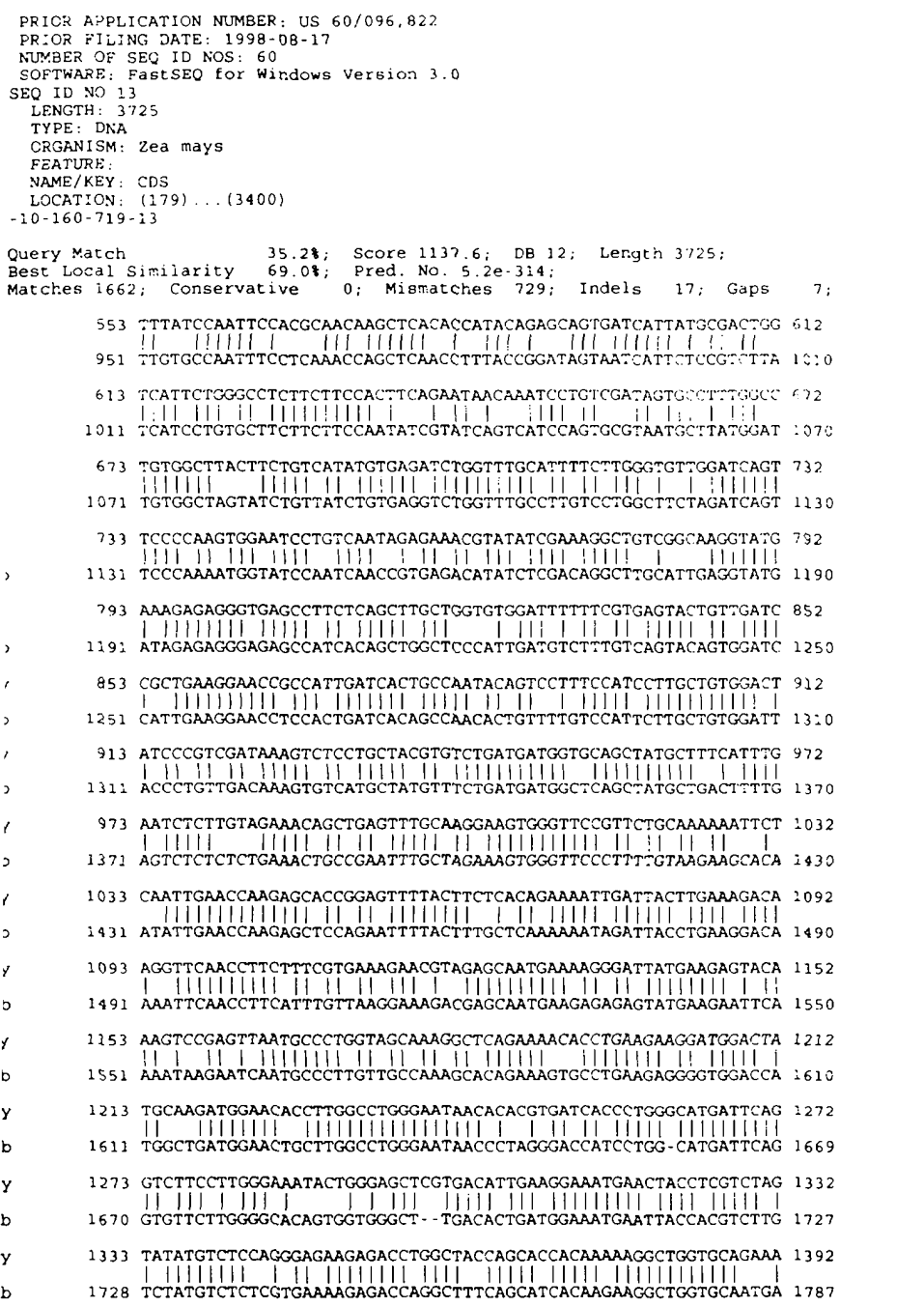
Qy       1033 CAATTGAACCAAGAGCACCGGAGTTTTACTTCTCACAGAAAATTGATTACTTGAAAGACA 1092
Db       1431 ATATTGAACCAAGAGCTCCAGAATTTTACTTTGCTCAAAAAATAGATTACCTGAAGGACA 1490
```

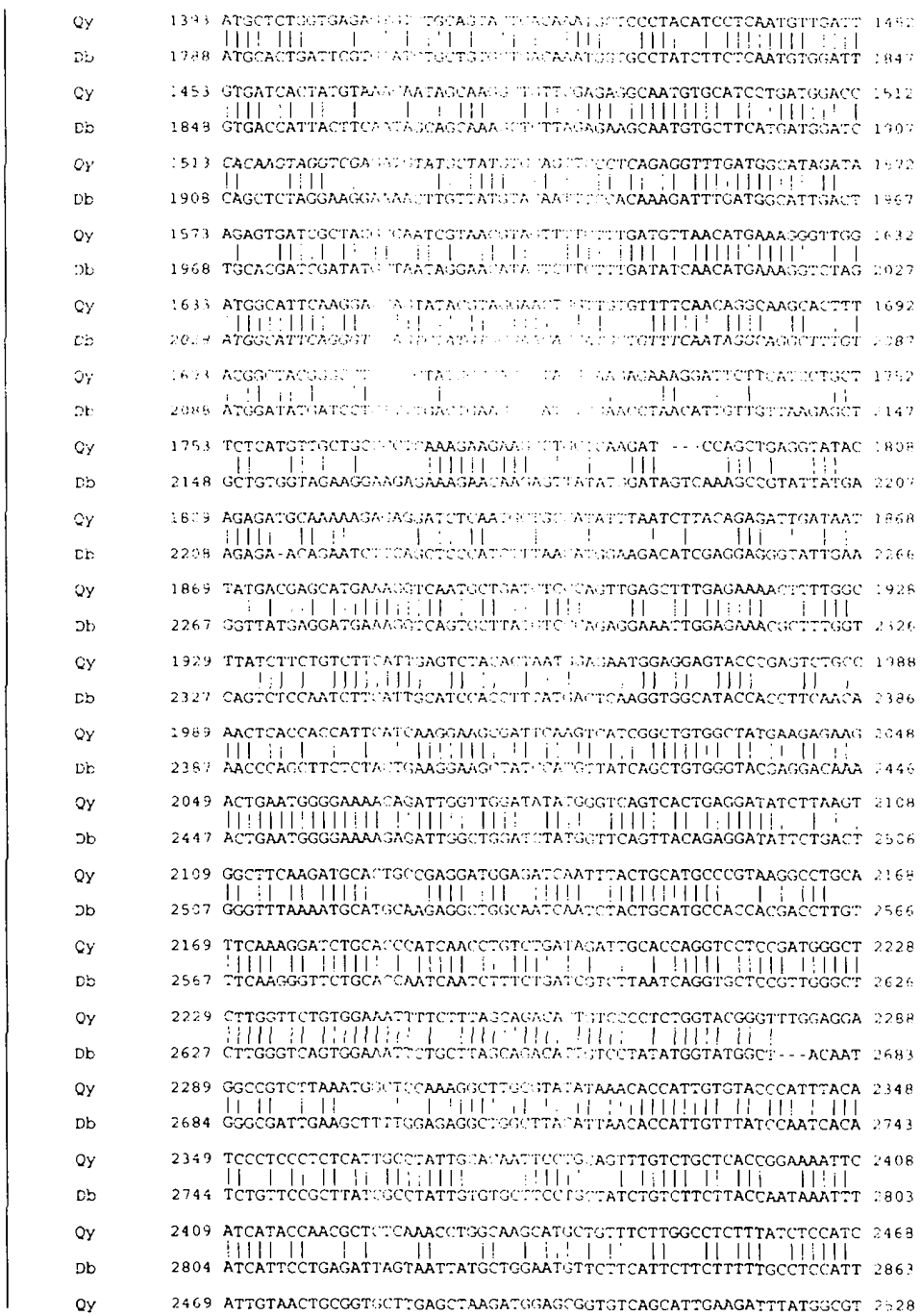


RESULT 11
US-10-160-719-13
; Sequence 13, Application US/10160719
; Publication No. US20030167528A1
; GENERAL INFORMATION:
; APPLICANT: Kanwarpal S. Dhugga
; APPLICANT: Timothy G. Helentjaris
; APPLICANT: Benjamin A. Bowen
; APPLICANT: Xun Wang
; TITLE OF INVENTION: Maize Cellulose Synthases and Uses
; TITLE OF INVENTION: Thereof
; FILE REFERENCE: 0864
; CURRENT APPLICATION NUMBER: US/10/160,719
; CURRENT FILING DATE: 2002-06-03
; PRIOR APPLICATION NUMBER: US/09/371,383A
; PRIOR FILING DATE: 1999-08-06

PRIOR APPLICATION NUMBER: US 60/096,822
PRIOR FILING DATE: 1998-08-17
NUMBER OF SEQ ID NOS: 60
SOFTWARE: PastSEQ for Windows Version 3.0
SEQ ID NO 13
LENGTH: 3725
TYPE: DNA
ORGANISM: Zea mays
FEATURE:
NAME/KEY: CDS
LOCATION: (179)...(3400)
-10-160-719-13

Query Match 35.2%; Score 1137.6; DB 12; Length 3725;
Best Local Similarity 69.0%; Pred. No. 5.2e-314;
Matches 1662; Conservative 0; Mismatches 729; Indels 17; Gaps 7;





```
Db       2864 TTCGCAACTGGTATATTGGAGCTCAGATGGAGTGGTGTTGGCATTGAAGATTGGTGGAGA 2923

Qy       2529 AATGAACAATTCTGGGTGATCGGAGGTGTTTCAGCCCATCTCTTTGCGGTCTTCCAGGGA 2588

Db       2924 AATGAGCAGTTTTGGGTTATTGGTGGCACCTCTGCCCATCTCTTCGCGGTGTTCCAGGGT 2983

Qy       2589 TTCTTAAAAATGTTGGCTGGCATCGATACGAACTTCACTGTCACAGCAAAAGCA---GCC 2645

Db       2984 CTGCTGAAAGTGTTGGCTGGGATTGATACCAACTTCACAGTTACCTCAAAGGCATCTGAT 3043

Qy       2646 GAAGATGCAGAATTTGGGGAGCTATATATGGTCAAGTGGACAACACTTTTGATTCCTCCA 2705

Db       3044 GAGGATGGCGACTTTGCTGAGCTATATGTGTTCAAGTGGACCAGTTTGCTCATCCCTCCG 3103

Qy       2706 ACCACACTTCTCATTATCAATATGTCGGG  TTGTGCTGGATTCTCTGATGCACTCAAC 2762

Db       3104 ACCACTGTTCTTGTCATTAACCTGGTGGAATGGTGGCAGGAATTTCGTATGCCATTAAC 3163

Qy       2763 AAAGGATATGAAGCATGGGGGCCTCTCTTTGGCAAGGTGTTCTTTGCTTTCTGGGTGATT 2822

Db       3164 AGCGGCTACCAATCCTGGGGTCCGCTCTTTGGAAAGCTGTTCTTCTCGATCTGGGTGATC 3223

Qy       2823 CTTCATCTCTATCCATTCCTTAAAGGTCTAATGGGTCGCCAAAACCTAACACCAACCATT 2882

Db       3224 CTCCATCTCTACCCCTTCCTCAAGGGTCTCATGGGCAGGCAGAACCGCACGCCAACAATC 3283

Qy       2883 GTTGTTCTCTGGTCAGTGCTGTTGGCCTCTGTCTTCTCTCTCGTTTGGGTCAAGATCAAT 2942

Db       3284 GTCATCGTTTGGTCCATCCTCTTGCGTCTATCTTCTCCTTGCTGTGGGTGAAGATCGAT 3343

Qy       2943 CCATTCGT 2950

Db       3344 CCTTTCAT 3351


RESULT 12
US-10-160-719-21
; Sequence 21, Application US/10160719
; Publication No. US20030167528A1
; GENERAL INFORMATION:
;  APPLICANT: Kanwarpal S. Dhugga
;  APPLICANT:  Timothy G. Helentjaris
;  APPLICANT:  Benjamin A.  Bowen
;  APPLICANT:  Xun Wang
;  TITLE OF INVENTION: Maize Cellulose Synthases and Uses
;  TITLE OF INVENTION:  Thereof
;  FILE REFERENCE: 0864
;  CURRENT APPLICATION NUMBER: US/10/160,719
;  CURRENT FILING DATE:  2002-06-04
;  PRIOR APPLICATION NUMBER: US/09/370,384A
;  PRIOR FILING DATE: 1999-08-06
;  PRIOR APPLICATION NUMBER: US 60/096,828
;  PRIOR FILING DATE: 1998-08-17
;  NUMBER OF SEQ ID NOS: 60
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 21
;  LENGTH: 3725
;  TYPE: DNA
;  ORGANISM: Zea mays
;  FEATURE:
;  NAME/KEY: CDS
;  LOCATION: (179)...(3400)
US-10-160-719-21

  Query Match           35.2%;  Score 1137.6;  DB 12;  Length 3725;
  Best Local Similarity   69.0%;  Pred. No. 5.2e-314;
  Matches 1662;  Conservative    0;  Mismatches  729;  Indels   17;  Gaps    7;

Qy        553 TTTATCCAATTCCACGCAACAAGCTCACACCATACAGAGCATGATCATTATGCGACTGG 612

Db        951 TTGTGCCAATTTCCTCAAACCAGCTCAACCCTTACCGGATACTAATCATTCTCCGTCTTA 1010

Qy        613 TCATTCTGGGCCTCTTCTTCCACTTCAGAATAACAAATCCTGTCGATAGTGCCTTTGGCC 672

Db       1011 TCATCCTGTGCTTCTTCTTCCAATATCGTATCAGTCATCCAGTGCGTAATGCTTATGGAT 1070

Qy        673 TGTGGCTTACTTCTGTCATATGTGAGATCTGGTTTGCATTTTCTTGGGTGTTGGATCAGT 732

Db       1071 TGTGGCTAGTATCTGTTATCTGTGAGGTCTGGTTTGCCTTGTCCTGGCTTCTAGATCAGT 1130

Qy        733 TCCCCAAGTGGAATCCTGTCAATAGAGAAACGTATATCGAAAGGCTGTCGGCAAGGTATG 792

Db       1131 TCCCAAAATGGTATCCAATCAACCGTGAGACATATCTCGACAGGCTTGCATTGAGGTATG 1190

Qy        793 AAAGAGAGGGTGAGCCTTCTCAGCTTGCTGGTGTGGATTTTTTCGTGAGTACTGTTGATC 852

Db       1191 ATAGAGAGGGAGAGCCATCACAGCTGGCTCCCATTGATGTCTTTGTCAGTACAGTGGATC 1250

Qy        853 CGCTGAAGGAACCGCCATTGATCACTGCCAATACAGTCCTTTCCATCCTTGCTGTGGACT 912

Db       1251 CATTGAAGGAACCTCCACTGATCACAGCCAACACTGTTTTGTCCATTCTTGCTGTGGATT 1310

Qy        913 ATCCCGTCGATAAAGTCTCCTGCTACGTGTCTGATGATGGTGCAGCTATGCTTTCATTTG 972

Db       1311 ACCCTGTTGACAAAGTGTCATGCTATGTTTCTGATGATGGCTCAGCTATGCTGACTTTTG 1370

Qy        973 AATCTCTTGTAGAAACAGCTGAGTTTGCAAGGAAGTGGGTTCCGTTCTGCAAAAAATTCT 1032

Db       1371 AGTCTCTCTCTGAAACTGCCGAATTTGCTAGAAAGTGGGTTCCCTTTTGTAAGAAGCACA 1430

Qy       1033 CAATTGAACCAAGAGCACCGGAGTTTTACTTCTCACAGAAAATTGATTACTTGAAAGACA 1092

Db       1431 ATATTGAACCAAGAGCTCCAGAATTTTACTTTGCTCAAAAAATAGATTACCTGAAGGACA 1490

Qy       1093 AGGTTCAACCTTCTTTCGTGAAAGAACGTAGAGCAATGAAAAGGGATTATGAAGAGTACA 1152

Db       1491 AAATTCAACCTTCATTTGTTAAGGAAAGACGAGCAATGAAGAGAGAGTATGAAGAATTCA 1550

Qy       1153 AAGTCCGAGTTAATGCCCTGGTAGCAAAGGCTCAGAAAACACCTGAAGAAGGATGGACTA 1212

Db       1551 AAATAAGAATCAATGCCCTTGTTGCCAAAGCACAGAAAGTGCCTGAAGAGGGGTGGACCA 1610

Qy       1213 TGCAAGATGGAACACCTTGGCCTGGGAATAACACACGTGATCACCCTGGGCATGATTCAG 1272

Db       1611 TGGCTGATGGAACTGCTTGGCCTGGGAATAACCCTAGGGACCATCCTGG-CATGATTCAG 1669

Qy       1273 GTCTTCCTTGGGAAATACTGGGAGCTCGTGACATTGAAGGAAATGAACTACCTCGTCTAG 1332

Db       1670 GTGTTCTTGGGGCACAGTGGTGGGCT--TGACACTGATGGAAATGAATTACCACGTCTTG 1727

Qy       1333 TATATGTCTCCAGGGAGAAGAGACCTGGCTACCAGCACCACAAAAAGGCTGGTGCAGAAA 1392

Db       1728 TCTATGTCTCTCGTGAAAAGAGACCAGGCTTTCAGCATCACAAGAAGGCTGGTGCAATGA 1787

Qy       1393 ATGCTCTGGTGAGAGTGTCTGCAGTACTCACAAATGCTCCCTACATCCTCAATGTTGATT 1452

Db       1788 ATGCACTGATTCGTGTATCTGCTGTGCTGACAAATGGTGCCTATCTTCTCAATGTGGATT 1847

Qy       1453 GTGATCACTATGTAAACAATAGCAAGGCTGTTCGAGAGGCAATGTGCATCCTGATGGACC 1512

Db       1848 GTGACCATTACTTCAATAGCAGCAAAGCTCTTAGAGAAGCAATGTGCTTCATGATGGATC 1907

Qy       1513 CACAAGTAGGTCGAGATGTATGCTATGTGCAGTTCCCTCAGAGGTTTGATGGCATAGATA 1572

Db       1908 CAGCTCTAGGAAGGAAAACTTGTTATGTACAATTTCCACAAAGATTTGATGGCATTGACT 1967

Qy       1573 AGAGTGATCGCTACGCCAATCGTAACGTAGTTTTCTTTGATGTTAACATGAAAGGGTTGG 1632

Db       1968 TGCACGATCGATATGCTAATAGGAACATAGTCTTCTTTGATATCAACATGAAAGGTCTAG 2027

Qy       1633 ATGGCATTCAAGGACCAGTATACGTAGGAACTGGTTGTGTTTTCAACAGGCAAGCACTTT 1692

Db       2028 ATGGCATTCAGGGTCCAGTCTATGTGGGAACAGGATGCTGTTTCAATAGGCAGGCTTTGT 2087

Qy       1693 ACGGCTACGGGCCTCCTTCTATGCCCAGCTTACGCAAGAGAAAGGATTCTTCATCCTGCT 1752
```

```
2088 ATGGATATGATCCTGTTTTGACTGAAGCTGATCTGGAACCTAACATTGTTGTTAAGAGCT 2147

1753 TCTCATGTTGCTGCCCCTCAAAGAAGAAGCCTGCTCAAGAT----CCAGCTGAGGTATAC 1808
2148 GCTGTGGTAGAAGGAAGAGAAAGAACAAGAGTTATATGGATAGTCAAAGCCGTATTATGA 2207

1809 AGAGATGCAAAAAGAGAGGATCTCAATGCTGCCATATTTAATCTTACAGAGATTGATAAT 1868
2208 AGAGA-ACAGAATCTTCAGCTCCCATCTTTAACATGGAAGACATCGAGGAGGGTATTGAA 2266

1869 TATGACGAGCATGAAAGGTCAATGCTGATCTCCCAGTTGAGCTTTGAGAAAACTTTTGGC 1928
2267 GGTTATGAGGATGAAAGGTCAGTGCTTATGTCCCAGAGGAAATTGGAGAAACGCTTTGGT 2326

1929 TTATCTTCTGTCTTCATTGAGTCTACACTAATGGAGAATGGAGGAGTACCCGAGTCTGCC 1988
2327 CAGTCTCCAATCTTCATTGCATCCACCTTTATGACTCAAGGTGGCATACCACCTTCAACA 2386

1989 AACTCACCACCATTCATCAAGGAAGCGATTCAAGTCATCGGCTGTGGCTATGAAGAGAAG 2048
2387 AACCCAGCTTCTCTACTGAAGGAAGCTATCCATGTTATCAGCTGTGGGTACGAGGACAAA 2446

2049 ACTGAATGGGGAAAACAGATTGGTTGGATATATGGGTCAGTCACTGAGGATATCTTAAGT 2108
2447 ACTGAATGGGGAAAAGAGATTGGCTGGATCTATGGTTCAGTTACAGAGGATATTCTGACT 2506

2109 GGCTTCAAGATGCACTGCCGAGGATGGAGATCAATTTACTGCATGCCCGTAAGGCCTGCA 2168
2507 GGGTTTAAAATGCATGCAAGAGGCTGGCAATCAATCTACTGCATGCCACCACGACCTTGT 2566

2169 TTCAAAGGATCTGCACCCATCAACCTGTCTGATAGATTGCACCAGGTCCTCCGATGGGCT 2228
2567 TTCAAGGGTTCTGCACCAATCAATCTTTCTGATCGTCTTAATCAGGTGCTCCGTTGGGCT 2626

2229 CTTGGTTCTGTGGAAATTTTCTTTAGCAGACACTGTCCCCTCTGGTACGGGTTTGGAGGA 2288
2627 CTTGGGTCAGTGGAAATTCTGCTTAGCAGACATTGTCCTATATGGTATGGCT---ACAAT 2683

2289 GGCCGTCTTAAATGGCTCCAAAGGCTTGCGTATATAAACACCATTGTGTACCCATTTACA 2348
2684 GGGCGATTGAAGCTTTTGGAGAGGCTGGCTTACATTAACACCATTGTTTATCCAATCACA 2743

2349 TCCCTCCCTCTCATTGCCTATTGCACAATTCCTGCAGTTTGTCTGCTCACCGGAAAATTC 2408
2744 TCTGTTCCGCTTATCGCCTATTGTGTGCTTCCTGCTATCTGTCTTCTTACCAATAAATTT 2803

2409 ATCATACCAACGCTCTCAAACCTGGCAAGCATGCTGTTTCTTGGCCTCTTTATCTCCATC 2468
2804 ATCATTCCTGAGATTAGTAATTATGCTGGAATGTTCTTCATTCTTCTTTTTGCCTCCATT 2863

2469 ATTGTAACTGCGGTGCTTGAGCTAAGATGGAGCGGTGTCAGCATTGAAGATTTATGGCGT 2528
2864 TTCGCAACTGGTATATTGGAGCTCAGATGGAGTGGTGTTGGCATTGAAGATTGGTGGAGA 2923

2529 AATGAACAATTCTGGGTGATCGGAGGTGTTTCAGCCCATCTCTTTGCGGTCTTCCAGGGA 2588
2924 AATGAGCAGTTTTGGGTTATTGGTGGCACCTCTGCCCATCTCTTCGCGGTGTTCCAGGGT 2983

2589 TTCTTAAAAATGTTGGCTGGCATCGATACGAACTTCACTGTCACAGCAAAAGCA---GCC 2645
2984 CTGCTGAAAGTGTTGGCTGGGATTGATACCAACTTCACAGTTACCTCAAAGGCATCTGAT 3043

2646 GAAGATGCAGAATTTGGGGAGCTATATATGGTCAAGTGGACAACACTTTTGATTCCTCCA 2705
3044 GAGGATGGCGACTTTGCTGAGCTATATGTGTTCAAGTGGACCAGTTTGCTCATCCCTCCG 3103

2706 ACCACACTTCTCATTATCAATATGTCGGG---TTGTGCTGGATTCTCTGATGCACTCAAC 2762
3104 ACCACTGTTCTTGTCATTAACCTGGTCGGAATGGTGGCAGGAATTTCGTATGCCATTAAC 3163

2763 AAAGGATATGAAGCATGGGGGCCTCTCTTTGGCAAGGTGTTCTTTGCTTTCTGGGTGATT 2822
```

```
Db       3164 AGCGGCTACCAATCCTGGGGTCCGTCTTTTGGAAAGCTGTTCTTCTCGATCTGGGTGATC 3223

Qy       2823 CTTCATCTCTATCCATTCCTTAAAGTTCTAATGGGTCGCCAAAACCTAACACCAACCATT 2882
Db       3224 CTCCATCTCTACCCTTCCTCAAGGTCTCATGGGCAGGCAGAACCGCACGCCAACAATC 3283

Qy       2883 GTTGTTCTCTGGTCAGTGCTGTTGGGCTCTGTCTTCTCTCTCGTTTGGGTCAAGATCAAT 2942
Db       3284 GTCATCGTTTGGTCCATCCTCCTTGCGTCTATCTTCTCCTTGCTGTGGGTGAAGATCGAT 3343

Qy       2943 CCATTCGT 2950
Db       3344 CCTTTCAT 3351


RESULT 13
US-10-160-719-41
; Sequence 41, Application US/10/160,719
; Publication No. US2003016[illegible]A1
; GENERAL INFORMATION:
;  APPLICANT: Kanwarpal S [illegible]
;  APPLICANT:  Timothy G. B[illegible]
;  APPLICANT:  Benjamin A [illegible]
;  APPLICANT:  Xun Wang
;  TITLE OF INVENTION: Maize Cellulose Synthases and Uses
;  TITLE OF INVENTION:  Thereof
;  FILE REFERENCE: 0864
;  CURRENT APPLICATION NUMBER: US/10/160,719
;  CURRENT FILING DATE:  2002-06-03
;  PRIOR APPLICATION NUMBER: US/09/371,483A
;  PRIOR FILING DATE: 1999-08-06
;  PRIOR APPLICATION NUMBER: US 60/096,822
;  PRIOR FILING DATE: 1998-08-17
;  NUMBER OF SEQ ID NOS: 60
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 41
;   LENGTH: 3725
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (179)...(3400)
US-10-160-719-41

  Query Match           35.2%;  Score 1137.6;  DB 12;  Length 3725;
  Best Local Similarity   69.0%;  Pred. No. 5.2e-314;
  Matches 1662;  Conservative    0;  Mismatches  729;  Indels   17;  Gaps    7;

Qy        553 TTTATCCAATTCCAGGCAACAAGCTCACACCATACAGAGCAGTGATCATTATGCGACTGG 612
Db        951 TTGTGCCAATTTCCTCAAACCAGCTCAACCTTCACCGGATAGTAATCATTCTCCGTCTTA 1010

Qy        613 TCATTCTGGGCCTCTTCTTCCACTTCAGAATAACAAATCCTGTCGATAGTGCCTTTGGCC 672
Db       1011 TCATCCTGTGCTTCTTCTTCCAATATCGTATCAGTCATCCAGTGCGTAATGCTTATGGAT 1070

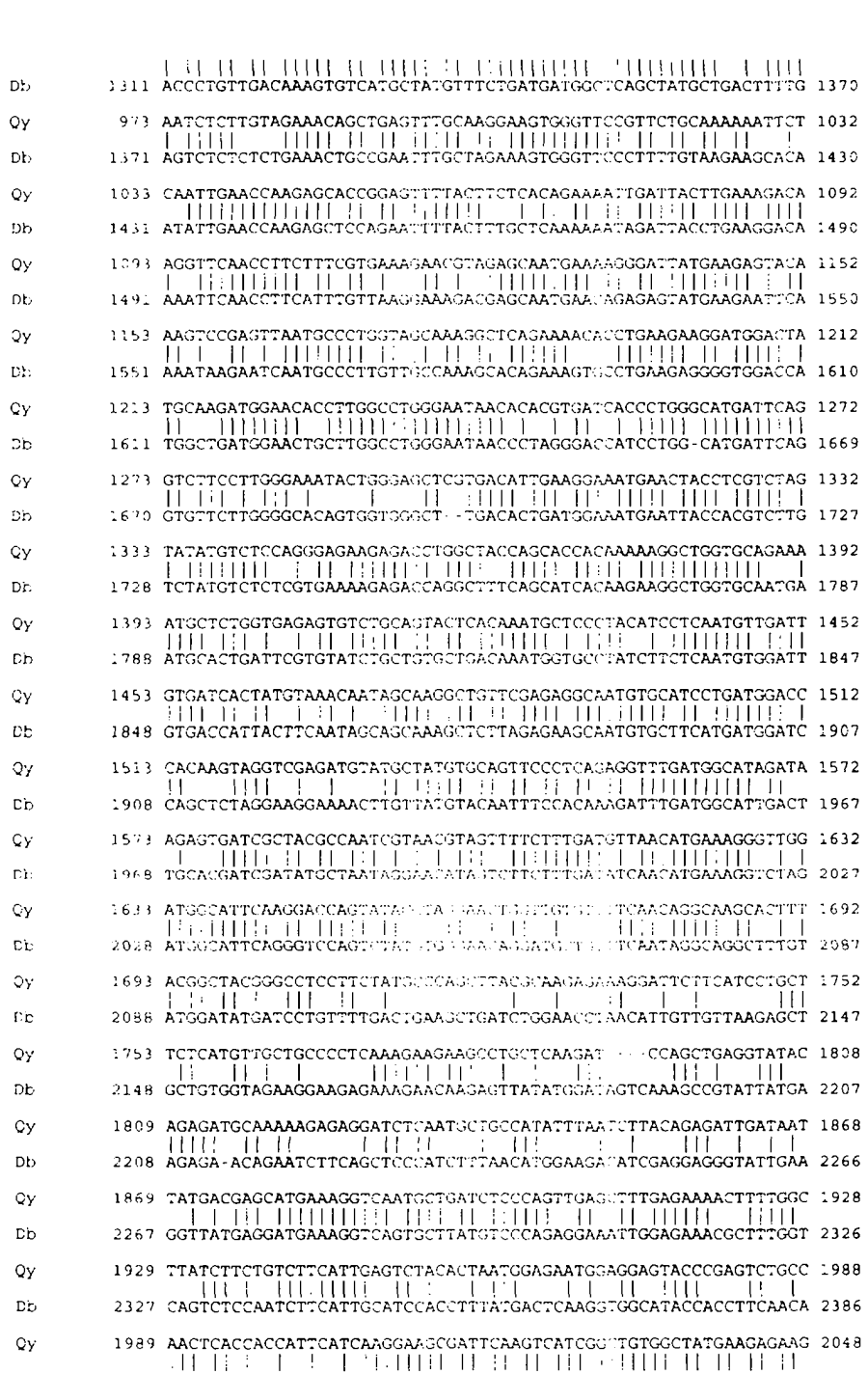
Qy        673 TGTGGCTTACTTCTGTCATATGTGAGATCTGGTTTGCATTTTCTTGGGTGTTGGATCAGT 732
Db       1071 TGTGGCTAGTATCTGTTATCTGTGAGGTCTGGTTTGCCTTGTCCTGGCTTCTAGATCAGT 1130

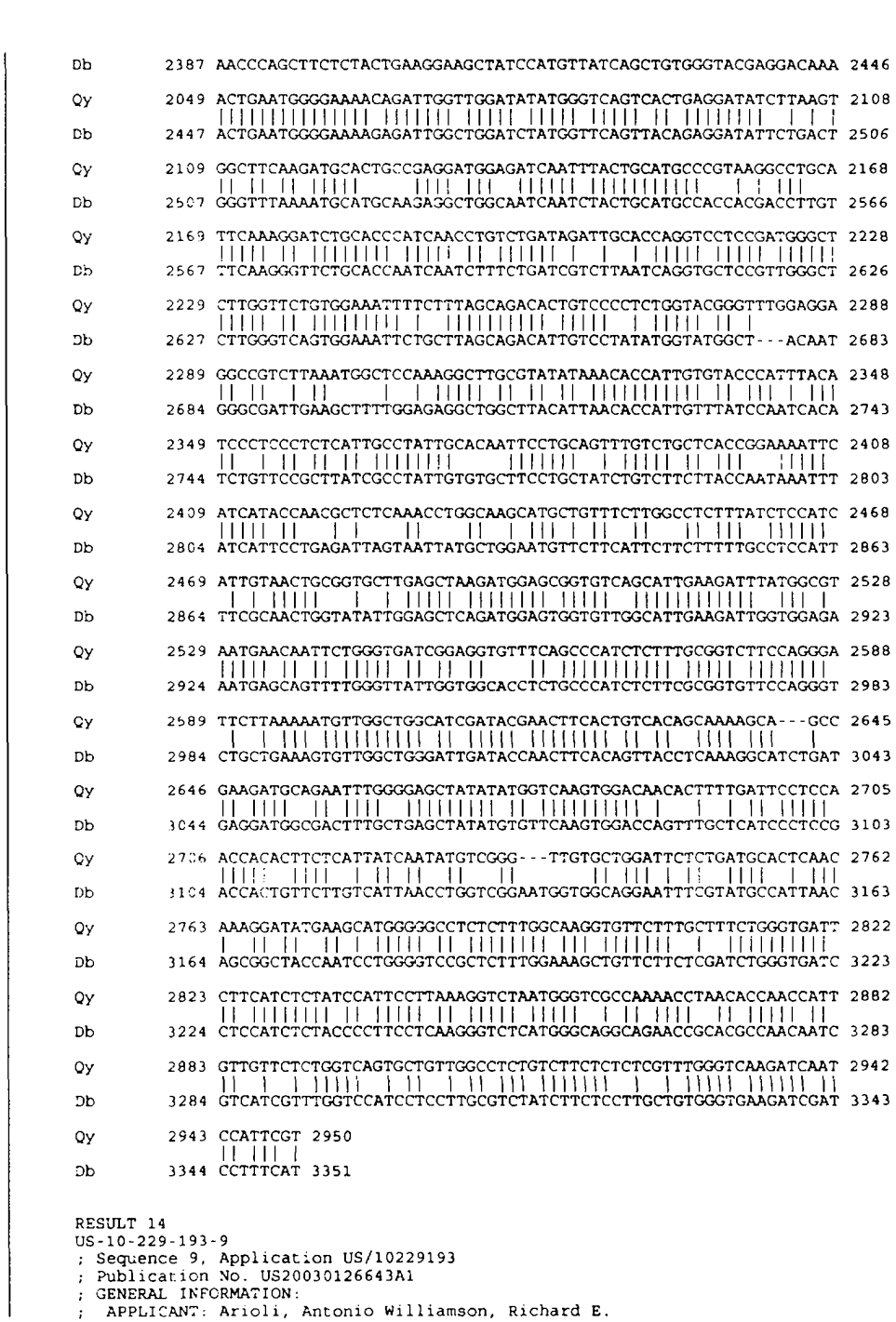
Qy        733 TCCCCAAGTGGAATCCTGTCAATAGAGAAACGTATATCGAAAGGCTGTCGGCAAGGTATG 792
Db       1131 TCCCAAAATGGTATCCAATCAACCGTGAGACATATCTCGACAGGCTTGCATTGAGGTATG 1190

Qy        793 AAAGAGAGGGTGAGCCTTCTCAGCTTGCTGGTGTGGATTTTTTCGTGAGTACTGTTGATC 852
Db       1191 ATAGAGAGGGAGAGCCATCACAGCTGGCTCCCATTGATGTCTTTGTCAGTACAGTGGATC 1250

Qy        853 CGCTGAAGGAACCGCCATTGATCACTGCCAATACAGTCCTTTCCATCCTTGCTGTGGACT 912
Db       1251 CATTGAAGGAACCTCCACTGATCACAGCCAACACTGTTTTGTCCATTCTTGCTGTGGATT 1310

Qy        913 ATCCCGTCGATAAAGTCTCCTGCTACGTGTCTGATGATGGTGCAGCTATGCTTTCATTTG 972
```





RESULT 14
US-10-229-193-9
; Sequence 9, Application US/10229193
; Publication No. US20030126643A1
; GENERAL INFORMATION:
; APPLICANT: Arioli, Antonio Williamson, Richard E.

APPLICANT: Betzner, Andreas S.
APPLICANT: Peng, Liangcai
TITLE OF INVENTION: Manipulation of Cellulose and/or Beta-1,4-Glucan
FILE REFERENCE: 96-98A
CURRENT APPLICATION NUMBER: US/10/229,193
CURRENT FILING DATE: 2002-08-26
PRIOR APPLICATION NUMBER: 09/221,013
PRIOR FILING DATE: 1998-12-23
PRIOR APPLICATION NUMBER: PCT/AU97/00402
PRIOR FILING DATE: 1997-06-24
PRIOR APPLICATION NUMBER: AU PO0699
PRIOR FILING DATE: 1996-06-27
NUMBER OF SEQ ID NOS: 36
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 9
LENGTH: 3614
TYPE: DNA
ORGANISM: Arabidopsis thaliana
FEATURE:
NAME/KEY: CDS
LOCATION: (217)..(3411)
-10-229-193-9

Query Match 35.1%; Score 1133.8; DB 14; Length 3614;
Best Local Similarity 68.1%; Pred. No. 5.6e-313;
Matches 1691; Conservative 0; Mismatches 762; Indels 30; Gaps 7;

```
          519 CCGTCTGCAGAGGCTTCGGAGCCGCTTTCAATTGTTTATCCAATTCCACGCAACAAGCTC 578
              |  |  ||  || ||   | |||  || ||||       | |||||||          | ||
          922 CTGAATGACGAAGCGAGGCAGCTTCTGTCAAGGAAAGTTTCAATTCCTTCATCACGGATC 981

          579 ACACCATACAGAGCAGTGATCATTATGCGACTGGTCATTCTGGGCCTCTTCTTCCACTTC 638
              |  || ||||||   || || ||  |||| || || || ||  | |||||||| || | |
          982 AATCCTTACAGAATGGTTATTATGCTGCGGCTTGTTATCCTTTGTCTCTTCTTGCATTAC 1041

          639 AGAATAACAAATCCTGTCGATAGTGCCTTTGGCCTGTGGCTTACTTCTGTCATATGTGAG 698
              | ||||||| || ||    | |||||||  || |||||    ||||| |||||||||
         1042 CGTATAACAAACCCAGTGCCAAATGCCTTTGCTCTATGGCTGGTCTCTGTGATATGTGAG 1101

          699 ATCTGGTTTGCATTTTCTTGGGTGTTGGATCAGTTCCCCAAGTGGAATCCTGTCAATAGA 758
              ||||||||||| || || ||| | |||||||||||| |||||||||  ||||| ||  |
         1102 ATCTGGTTTGCCTTATCCTGGATTTTGGATCAGTTTCCCAAGTGGTTTCCTGTGAACCGT 1161

          759 GAAACGTATATCGAAAGGCTGTCGGCAAGGTATGAAAGAGAGGGTGAGCCTTCTCAGCTT 818
              ||||| ||  |||| |||||  |   ||| |||||  | || |||||||| || | |
         1162 GAAACCTACCTCGACAGGCTTGCTTTAAGATATGATCGTGAAGGTGAGCCATCACAGTTA 1221

          819 GCTGGTGTGGATTTTTTCGTGAGTACTGTTGATCCGCTGAAGGAACCGCCATTGATCACT 878
              |||| ||| ||  |||||||||||||||||||| ||  |||||| || ||  |  | ||
         1222 GCTGCTGTTGACATTTTCGTGAGTACTGTTGACCCCTTGAAGGAGCCACCCCTTGTGACA 1281

          879 GCCAATACAGTCCTTTCCATCCTTGCTGTGGACTATCCCGTCGATAAAGTCTCCTGCTAC 938
              ||||| ||||| || || || || ||||| ||||| || || || || || |||| ||
         1282 GCCAACACAGTGCTCTCTATTCTGGCTGTTGACTACCCAGTTGACAAGGTGTCCTGTTAT 1341

          939 GTGTCTGATGATGGTGCAGCTATGCTTTCATTTGAATCTCTTGTAGAAACAGCTGAGTTT 998
              || |||||||||||||| |||||| | |||||||||| |||| |||||| | ||||||
         1342 GTTTCTGATGATGGTGCTGCTATGTTATCATTGAATCACTTGCAGAAACATCAGAGTTT 1401

          999 GCAAGGAAGTGGGTTCCGTTCTGCAAAAAATTCTCAATTGAACCAAGAGCACCGGAGTTT 1058
              ||  | || ||||| || || ||||| ||||      || || ||  | ||||| || |
         1402 GCTCGTAAATGGGTACCATTTTGCAAGAAATATAGCATAGAGCCTCGTGCACCAGAATGG 1461

         1059 TACTTCTCACAGAAAATTGATTACTTGAAAGACAAGGTTCAACCTTCTTTCGTGAAAGAA 1118
              |||||  |  |||||| |||||||||| || || |||||  | || || || |||||
         1462 TACTTTGCTGCGAAAATAGATTACTTGAAGGATAAAGTTCAGACATCATTTGTCAAAGAT 1521

         1119 CGTAGAGCAATGAAAAGGGATTATGAAGAGTACAAAGTCCGAGTTAATGCCCTGGTAGCA 1178
              |||||||| ||||| ||||| |||| || |  ||| ||||| | ||||| || ||  |
         1522 CGTAGAGCTATGAAGAGGGAATATGAGGAATTTAAAATCCGAATCAATGCACTTGTTTCC 1581

         1179 AAGGCTCAGAAAACACCTGAAGAAGGATGGACTATGCAAGATGGAACACCTTGGCCTGGG 1238
              || || |  |||  |||||||||||| ||| |||||||||||| ||||| ||||||||
```

```
Db       1582 AAAGCCCTAAAATGTTCTGAAGAAGGGTGTGTTATGCAAGATGGCACACCGTGGCCTGGA 1641

Qy       1239 AATAACACACGTGATCACCCTGGGTATGATTCAGGTCTTCCTTGGGAAATACTGGGAGCT 1298
              ||||| ||| | || |: || || || | | | ||||| | ||| || |  ;| |
Db       1642 AATAATACAGGGGACCATCC-AGGAATGATCCAGGTCTTCTTAGGGCAAAA--TGGTGGA 1698

Qy       1299 CGTGACATTGAAGGAAATGAACTAGCTCGTCTAGTATATGTCTCCAGGGAGAAGAGACCT 1358
              | ||;   :| ||  |||| || ;  || | || | ||| ||  | || ||| ||||
Db       1699 CTTGATGCAGAGGGCAATGAGCTCCCGCGTTTGGTATATGTTTCTCGAGAAAAGCGACCA 1758

Qy       1359 GGCTACCAGCACCACAAAAAGGCTGGTGCAGAAAATGCTCTGGTGAGAGTGTCTGCAGTA 1418
              || | |||||||||||:|||||||||||:  :|||| |||||||||||| || ||| :
Db       1759 GGATTCCAGCACCACAAAAAGGCTGGTGTTATGAATGCACTGGTGAGAGTTTCAGCAGTT 1818

Qy       1419 CTCACAAATGCTCCTTACATCCTCAATGTTGATTGTGATCACTATGTAAACAATAGCAAG 1478
              || || |||  ||  ||| |  | |:|||||||||| ||  |||| || ||||
Db       1819 CTTACCAATGGACCTTTCATCTTGAATCTTGATTGTGATCATTACATAAATAACAGCAAA 1878

Qy       1479 GCTGTTCGAGAGGCAAGTGCATCCTTAAGGACCACAAGTAGGTCGAGATGTATGCTAT 1538
              || |  |||| |:| ||  |||| | | ||  |  ||   |  ||  ||| ||
Db       1879 GCCTTAAGAGAAGTCAGCTGTTCTTGAAGAAACCTTGGGAAGCAAGTTTGTTAT 1938

Qy       1539 GTGCAGTTCCCTCAAGTTTTGATCAAAGAAGAGTGATCGCTACGCCAATCGTAAC 1598
              || |,|||  | |:  | |:||||  ||  |||  ||| | || ||  ||||:|||
Db       1939 GTTCAGTTCCCACAAAGATTTGATGGTATCGACAAGAACGATAGATATGCTAATCGTAAT 1998

Qy       1599 GTAGTTTTCTTTGATGTTAACATGAAAGGGTTGGATGGCATTCAAGGACCAGTATACGTA 1658
              ||  |||||||||| |||| ||| ||  |||  ||| |||||||||| ,|||| ||,
Db       1999 ACCGTGTTCTTTGATATTAACTTGAGAGGTTTAGATGGGATTCAAGGACCTGTATATGTC 2058

Qy       1659 GGAACTGGTTGTGTTTTCAACAGCCAAGCACTTTACGGCTACGGGCCTCCTTCTATGCCC 1718
              |||||||| |||| |||| |||   |||  | ||,|  || |  ||||   | ||
Db       2059 GGAACTGGATGTGTTTTCAACAGAACAGCATTATACGGTTATGAACCTC---CAATAAAA 2115

Qy       1719 AGCTTACGCAAGAGAAAGGATTCTTCATCTTGTTTCTCATGTTGCTGCCCCTCAAAGAAG 1778
              |  ||||      ||  |  |||  || | ||||  |||,
Db       2116 GTAAAACACAAGAAGCCAAGTCTTTTATCTAAGTTCTGTGGTGGATCAAGAAAGAAGAAT 2175

Qy       1779 AAGCCTGCTCAAGATCCAGCTGAGGTATACAGATATGCAAAAAGAGAGGATCTCAATGCT 1838
              |||  |  |  ||  | | |  |  |||  ||  |  |  |  |  |  |  | | |
Db       2176 TCCAAAGCTAAGAAAGAGTCGGACAAAAAGAAATCAGGCAGGCATACTGACTCAACTGTT 2235

Qy       1839 GCCATATTTAATCTTACAGAGATTGATAA- - - - - - - - - -TTATGACGAGCATGAA 1883
              |  |||| |, ||    || :, ||                    , |  ,|  |,,|
Db       2236 CCTGTATTCAACCTCGATGACATAGAAGAGGGAGTTGAAGGTGCTGGTTTTGATGATGAA 2295

Qy       1884 AGGTCAATGCTGATCTCCCAGTTGAGCTTTGAGAAAACTTTTGGCTTATCTTCTGTCTTC 1943
              | | |  |  | || ||  ;  | |||  |||||    |||||    |||  |||  ||
Db       2296 AAGGCGCTCTTAATGTCGCAAATGAGCCTGGAGAAGCGATTTGGACAGTCTGCTGTTTTT 2355

Qy       1944 ATTGAGTCTACACTAATGGAGAATGGAGGAGTACCCGAGTCTGCCAACTCACCACCATTC 2003
              |||  ||||| |||.|||| |||||  .|  || .|   ||  |  ||  |  .||
Db       2356 GTTGCTTCTACCCTAATGGAAAATGGTGGTGTTCCTCCTTCAGCAACTCCAGAAAACTTT 2415

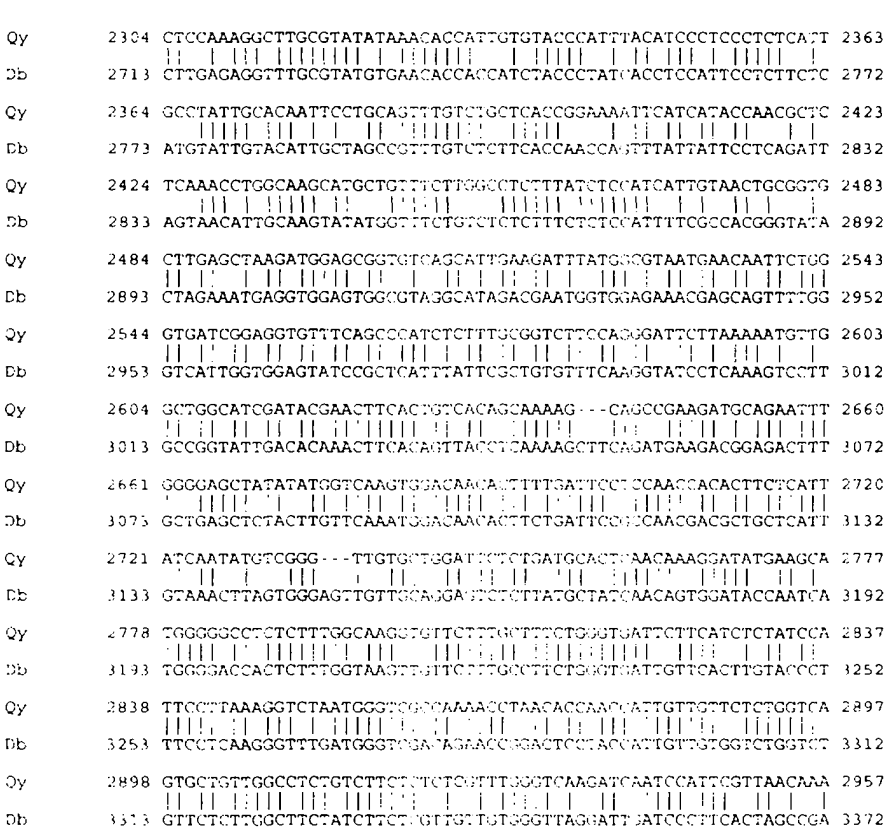
Qy       2004 ATCAAGGAAGCGATTCAAGTCATCGGCTGTGGCTATGAAGAGAAGACTGAATGGGGAAAA 2063
              |:|| |: || ;; || |||||  | |;|| || | || ||| | |. ||||||
Db       2416 CTCAAAGAGGCTATTCATGTCATTAGTTGTGGTTATGAGGATAAGTCAGATTGGGGAATG 2475

Qy       2064 CAGATTGGTTGGATATATGGGTCAGTCACTGAGGATATCTTAAGTGGCTTCAAGATGCAC 2123
              |||||||  |||| |||| ||  |  ||  |  |:||  | ||| ||||| ||||!
Db       2476 GAGATTGGATGGATTATGGTTCTGTGACAGAAGATATTCTGACTGGGTTCAAAATGCAT 2535

Qy       2124 TGCCGAGGATGGAGATCAATTTACTGCATGCCCGTAAGGCCTGCATTCAAAGGATCTGCA 2183
              ||| |||||| |.|| ||||||||||||||:          || || |||:| || :||||
Db       2536 GCCCGTGGATGGCGATCCATTTACTGCATGCCTAAGCTTCCAGCTTTCAAGGGTTCTGCT 2595

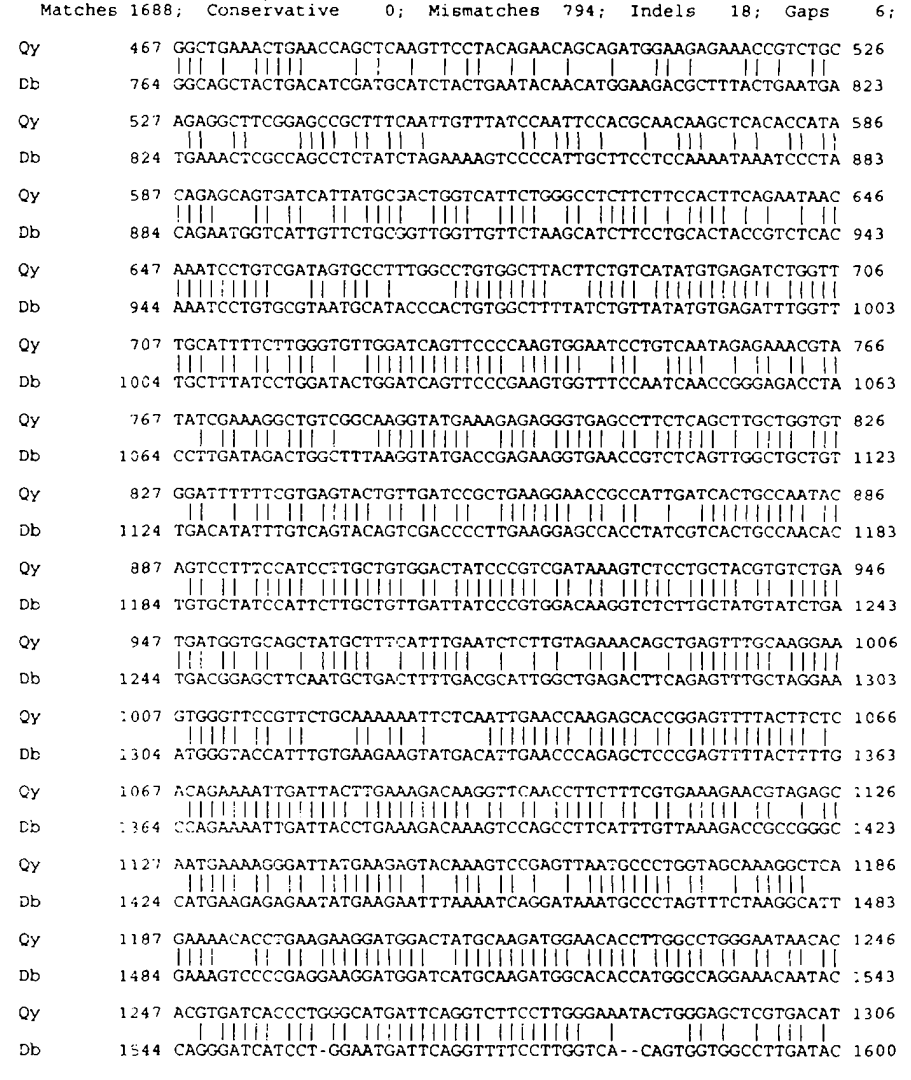
Qy       2184 CCCATCAACCTGTCTGATAGATTGCACCAGGTCCTCCGATGGGCTCTTGGTTCTGTGGAA 2243
              || ||||| || |, ||' |  |: |||| || :|  | ||||||  : |||| || ||
Db       2596 CCTATCAATCTTTCAGATCGTCTGAACCAAGTGCTGAGGTGGGCTTTAGGTTCAGTTGAG 2655

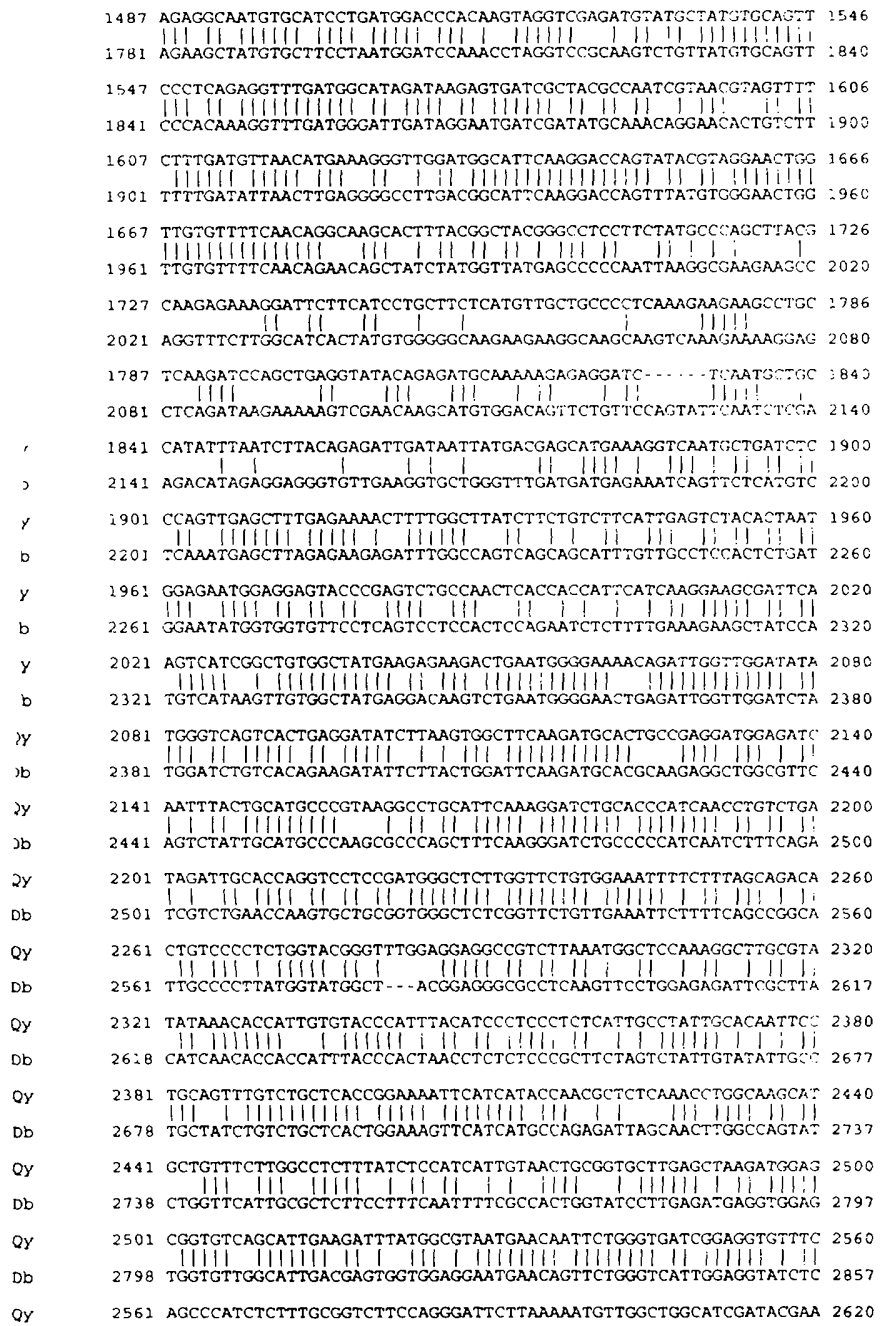
Qy       2244 ATTTTCTTTAGCAGACACTGTCCCCTCTGGTACGGGTTTGGAGGAGGCCGTCTTAAATGG 2303
              ||| |||| ,|  | || ||||  | |||||  |  |  || | ||| |||!
Db       2656 ATTCTCTTCAGTCGGCATTGTCCTATATGGTATGG- --TTACAATGGGAGGCTAAAATTT 2712
```

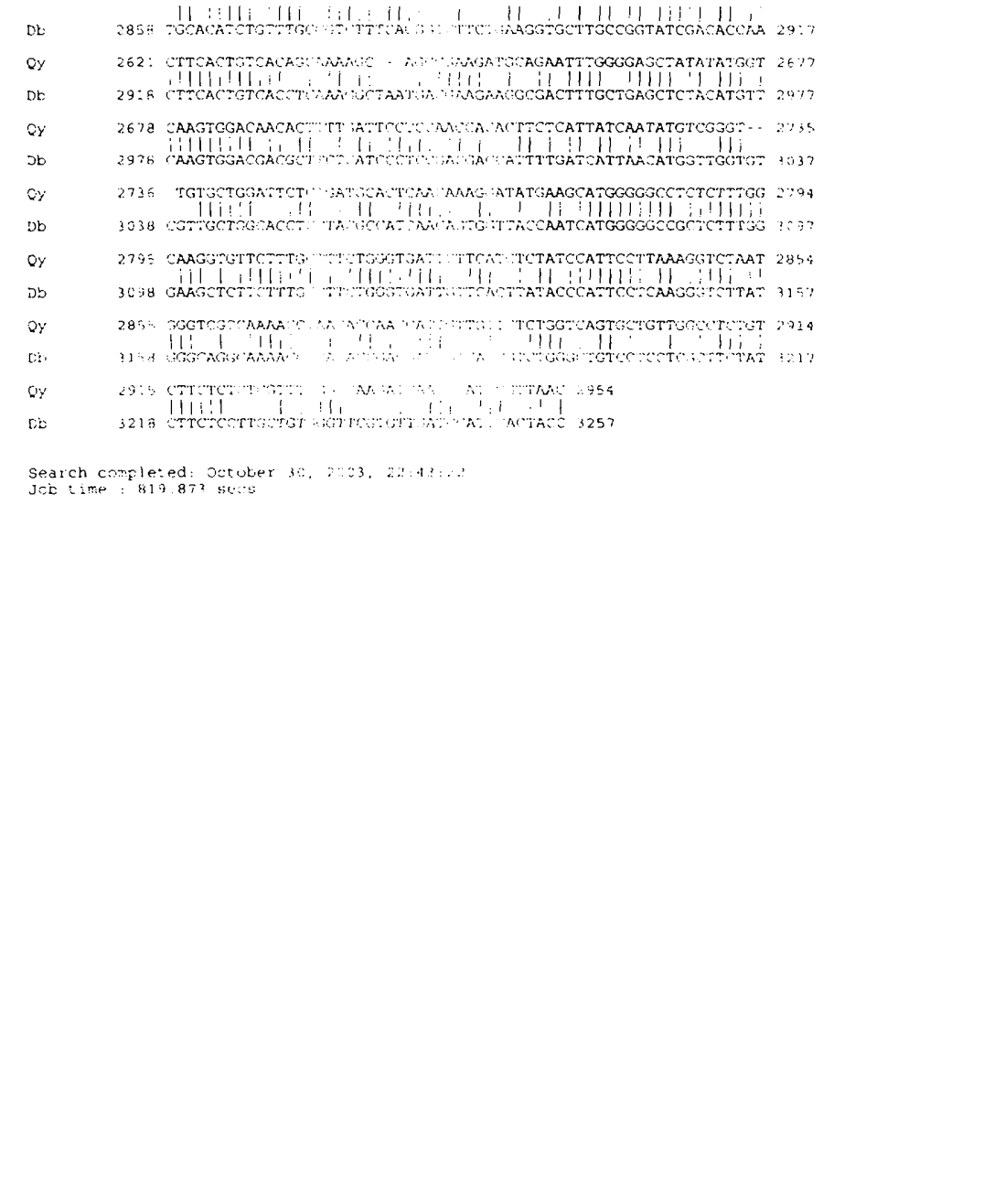


RESULT 15
US-09-900-237-29
; Sequence 29, Application US/09900237
; Patent No. US20020120124A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Stephen
; TITLE OF INVENTION: Plant Cellulose Synthases
; FILE REFERENCE: BB1170 US CIP
; CURRENT APPLICATION NUMBER: US/09/900,237
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: 60/092,844
; PRIOR FILING DATE: 1998-07-14
; PRIOR APPLICATION NUMBER: PCT/US99/15831
; PRIOR FILING DATE: 1999-07-13
; PRIOR APPLICATION NUMBER: 09/720383
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 29
; LENGTH: 3626
; TYPE: DNA
; ORGANISM: Triticum aestivum
US-09-900-237-29

Query Match 35.1%; Score 1133.6; DB 10; Length 3626;
Best Local Similarity 67.5%; Pred. No. 6.4e-313;
Matches 1688; Conservative 0; Mismatches 794; Indels 18; Gaps 6;







Search completed: October 30, 2003, 22:48:12
Job time : 819.873 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

CM protein: nucleic search, using frame_plus_p2n model

Run on: October 31, 2003, 09:33:04 ; Search time 332.263 Seconds
(without alignments)
4096.305 Million cell updates/sec

Title: US-09-980-043-4
Perfect score: 324
Sequence: 1 MMESGAPICHTCGEQVGHDA.....KSCFEPEIKEGRKVCLRCGS 56

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame_p2n.model -DEV=xlh
-Q /cgn2_1/USPTO_spool/US09980043/runat_29102003_163853_21871/app_query.fasta_1.1358
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09980043_@CGN_1_1_4125_@runat_29102003_163853_21871 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=-1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP 6 -DELEXT=7

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	317	97.8	434	13	BU894130	BU894130 X003C01 P
2	246	75.9	582	9	AV830247	AV830247 AV830247
3	236	72.8	502	12	BJ559283	BJ559283 BJ559283
4	236	72.8	506	12	BJ564652	BJ564652 BJ564652
5	236	72.8	511	12	BJ554190	BJ554190 BJ554190
6	234.5	72.4	378	10	BG135016	BG135016 EST467908
7	231	71.3	661	10	BG453374	BG453374 NF091C07L
8	183	56.5	515	10	BE326104	BE326104 NF086F04S
9	183	56.5	558	14	CA919485	CA919485 EST637203
10	175	54.0	500	10	BE203672	BE203672 EST396348
11	175	54.0	533	9	AW585621	AW585621 EST317244
12	175	54.0	537	13	BQ141412	BQ141412 NF019C12P
c 13	175	54.0	584	9	AJ548388	AJ548388 AJ548388
14	175	54.0	639	9	AW693843	AW693843 NF069F10S
15	175	54.0	643	12	BI307877	BI307877 EST529287
16	175	54.0	664	10	BF642644	BF642644 NF063H03I
17	175	54.0	666	9	AW693075	AW693075 NF059G10S
18	175	54.0	694	9	AW584983	AW584983 N211152e
19	175	54.0	834	14	CA917317	CA917317 EST641464
20	172	53.1	431	9	AW693491	AW693491 NF065E09S
21	172	53.1	481	12	BJ564247	BJ564247 BJ564247
22	172	53.1	528	12	BJ554611	BJ554611 BJ554611
23	172	53.1	579	12	BJ558386	BJ558386 BJ558386
c 24	169.5	52.3	512	9	AL368412	AL368412 MtBA24C11
25	168	51.9	374	9	AL370351	AL370351 MtBA37C05
26	168	51.9	683	13	BX255611	BX255611 BX255611
27	165	50.9	621	9	AU236888	AU236888 AU236888
c 28	163	50.3	569	9	AJ548175	AJ548175 AJ548175
29	163	50.3	721	14	CA498439	CA498439 WHE3243_D
30	162	50.0	382	10	BG053825	BG053825 RHIZ2_10_
31	162	50.0	906	10	BE216853	BE216853 HV_CEB001
32	162	50.0	3898	11	AY110415	AY110415 Zea mays
33	161	49.7	188	14	CD024532	CD024532 NXRV_044_
34	161	49.7	207	10	BE582127	BE582127 NXCI_029_
35	161	49.7	376	10	BF516943	BF516943 NXSI_007_
36	161	49.7	473	10	BG103336	BG103336 RHIZ2_18_
37	161	49.7	563	12	BM892555	BM892555 sam61f07.
38	161	49.7	585	12	BM893295	BM893295 sam54c08.
39	161	49.7	589	13	BQ700740	BQ700740 NXRV110_F
40	161	49.7	632	12	BM378215	BM378215 MEST275-B
41	161	49.7	646	12	BM378268	BM378268 MEST368-A
42	161	49.7	648	12	BM378257	BM378257 MEST333-H
43	161	49.7	668	12	BM492221	BM492221 NXRV_022_
44	160	49.4	398	12	BM736574	BM736574 952038C07
45	160	49.4	409	12	BM660669	BM660669 952040F08

ALIGNMENTS

RESULT 1
BU894130
LOCUS BU894130 434 bp mRNA linear EST 17-OCT-2002
DEFINITION X003C01 Populus wood cDNA library Populus tremula x Populus tremuloides cDNA 5 prime, mRNA sequence.
ACCESSION BU894130
VERSION BU894130.1 GI:24105195
KEYWORDS EST.
SOURCE Populus tremula x Populus tremuloides
ORGANISM Populus tremula x Populus tremuloides
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids, eurosids I; Malpighiales; Salicaceae; Populus.

```
FERENCE   1  (bases 1 to 434)
AUTHORS   Unneberg,P., Bhalerao,R.R., Jansson,S. and Sterky,F.
TITLE     The poplar tree transcriptome: Analysis of expressed sequence tags
          from multiple libraries
JOURNAL   Unpublished
OMMENT    Contact: BHALERAO RUPALI R.
          Umea Plant Science Center
          Department of Plant Physiology
          University of Umea, 901 87 Umea, Sweden
          Tel: +46 90 786 5279
          Fax: +46 90 786 6676
          Email: rupali.bhalerao@plantphys.umu.se.
ATURES              Location/Qualifiers
    source          1. .434
                    /organism="Populus tremula x Populus tremuloides"
                    /mol_type="mRNA"
                    /db_xref="taxon:47664"
                    /tissue_type="wood"
                    /clone_lib="Populus wood cDNA library"
ASE COUNT      132 a     66 c    125 g    111 t
RIGIN

lignment Scores:
red. No.:              8.46e-31       Length:         434
core:                  317.00         Matches:        54
ercent Similarity:     98.21%         Conservative:   1
est Local Similarity:  96.43%         Mismatches:     1
uery Match:            97.84%         Indels:         0
B:                     13             Gaps:           0

S-09-980 043-4 (1-56) x BU894130 (1-434)

y          1 MetMetGluSerGlyAlaProIleCysHisThrCysGlyGluGlnValGlyHisAspAla 20
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
o         62 ATGATGGAATCTGGGGCTCCTATATGCCATACCTGTGGTGAACAGGTGGGGCATGATGCA 121

y         21 AsnGlyGluLeuPheValAlaCysHisGluCysSerTyrProMetCysLysSerCysPhe 40
             ||||||||||||||||  ||||||||||||||:::|||||||||||||||||||||||||
o        122 AATGGGGAGCTATTTGCTGCTTGCCATGAGTGTAACTATCCCATGTGCAAGTCTTGTTTC 181

y         41 GluPheGluIleLysGluGlyArgLysValCysLeuArgCysGlySer 56
             ||||||||||||||||||||||||||||||||||||||||||||||||
o        182 GAGTTTGAAATCAAAGAGGGTCGGAAAGTTTGCTTGCGGTGTGGCTCG 229

ESULT 2
V830247
OCUS        AV830247                582 bp    mRNA    linear   EST 01-APR-2002
EFINITION   AV830247 RAFL9 Arabidopsis thaliana cDNA clone RAFL09-65-M12 5',
            mRNA sequence.
CCESSION    AV830247
ERSION      AV830247.1  GI:19872307
EYWORDS     EST.
OURCE       Arabidopsis thaliana (thale cress)
 ORGANISM   Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
EFERENCE    1  (bases 1 to 582)
 AUTHORS    Seki,M., Narusaka,M., Ishida,J., Kamiya,A., Satou,M., Nakajima,M.,
            Oono,Y., Sakurai,T., Carninci,P., Kawai,J., Itoh,M., Ishii,Y.,
            Arakawa,T., Shibata,K., Shinagawa,A., Muramatsu,M., Hayashizaki,Y.
            and Shinozaki,K.
 TITLE      Large scale analysis of Arabidopsis full-length cDNA (2002b)
 JOURNAL    Unpublished
OMMENT      Contact: Motoaki Seki
            Plant Functional Genomics Research Group
            RIKEN Genomic Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: 81-298-36-4359
            Fax: 81-298-36-9060
            Email: mseki@rtc.riken.go.jp
            An Arabidopsis full-length cDNA library was constructed essentially
            as reported previously (Seki et al., 1998).cDNA cleaved with BamHI
            and XhoI was ligated to modified Lambda FLC-1 vector (Carninci et
            al., submitted for publication) digested with BamHI and SalI. This
            clone is in a modified pBluescript vector. Please visit our web
            site (http://www.gsc.riken.go.jp/e/plant/index_e.html) for further
            details.
FEATURES              Location/Qualifiers
     source           1. .582
                      /organism="Arabidopsis thaliana"
                      /mol_type="mRNA"
                      /db_xref="taxon:3702"
                      /clone="RAFL09-65-M12"
                      /dev_stage="plants at various developmental stages from
                      germination to mature seeds"
                      /lab_host="DH10B"
                      /clone_lib="RAFL9"
                      /note="Site_1: BamHI; Site_2: SalI; subjected to
                      dehydration (1, 2, 5, 10, 24 hr) and cold (1, 2, 5, 10, 24
                      hr) treatments"
BASE COUNT      198 a    114 c    126 g    [illegible] t      5 others
ORIGIN

Alignment Scores:
Pred. No.:                1.7e-11        Length:         582
Score:                    246.00         Matches:        38
Percent Similarity:       89.29%         Conservative:   12
Best Local Similarity:    67.86%         Mismatches:     6
Query Match:              75.93%         Indels:         0
DB:                       9              Gaps:           0

US-09-980-043 4 (1-56) x AV830247 (1-582)

Qy          1 MetMetGluSerGlyAlaProIleCysHisThrCysGlyGluGlnValGlyHisAspAla 20
              ||||||||||||'   :::|||:||||:::|:: ||||||||:::::|||      :::
Db         69 ATGATGGAGTCTAGGTCTCCCATCTGCAACACTTGTGGTGAAGAGATTGGTGTAAAATCA 128

Qy         21 AsnGlyGluLeuPheValAlaCysHisGluCysSerTyrProMetCysLysSerCysPhe 40
              |||||||||   ||||| ||| ||||:|:|||| |:::|:|:::|:|||||:::|:|
Db        129 AACGGAGAGTTCTTTGTGGCTTGTCATGAGTGTAGTTTCCCAATCTGCAAAGCTTGTCTT 188

Qy         41 GluPheGluIleLysGluGlyArgLysValCysLeuArgCysGlySer 56
              |||:::|||   |||||||||||:::::::|:|:| ||||:||||:::
Db        189 GAGTATGAATTCAAAGAAGGTCGAAGAATTTGCTTGCGTTGCGGCAAT 236

RESULT 3
BJ559283
LOCUS       BJ559283                502 bp    mRNA    linear   EST 18-DEC-2002
DEFINITION  BJ559283 Ipomoea nil mixture of flower and flower bud Ipomoea nil
            cDNA clone jm28c14 5', mRNA sequence
ACCESSION   BJ559283
VERSION     BJ559283.1  GI:27241103
KEYWORDS    EST.
SOURCE      Ipomoea nil (Japanese morning glory)
  ORGANISM  Ipomoea nil
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            asterids; lamiids; Solanales; Convolvulaceae; Ipomoea.
REFERENCE   1  (bases 1 to 502)
  AUTHORS   Hoshino,A., Seki,M., Shin-i,T., Carninci,P., Kamiya,A., Shiraki,T.,
            Nitasaka,E., Shinozaki,K., Hayashizaki,Y., Kohara,Y. and Iida,S.
  TITLE     ESTs of Japanese morning glory
  JOURNAL   Unpublished
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES              Location/Qualifiers
     source           1. .502
                      /organism="Ipomoea nil"
                      /mol_type="mRNA"
                      /cultivar="Tokyo-kokei standard"
```

```
                     /db_xref="taxon:35883"
                     /clone="jm28c14"
                     /tissue_type="mixture of flower and flower bud"
                     /clone_lib="Ipomoea nil mixture of flower and flower bud"
BASE COUNT      161 a    100 c    140 g    101 t
ORIGIN

Alignment Scores:
Pred. No.:              2.68e-20        Length:         502
Score:                  236.00          Matches:        36
Percent Similarity:     83.93%          Conservative:   11
Best Local Similarity:  64.29%          Mismatches:     9
Query Match:            72.84%          Indels:         0
DB:                     12              Gaps:           0

US-09-980-043-4 (1-56) x BJ559283 (1-502)

Qy          1 MetMetGluSerGlyAlaProIleCysHisThrCysGlyGluGlnValGlyHisAspAla 20
              ||||||:::||||||   | |   :|:      ||||:|||:|:||:::|||      :::
Db         20 ATGATGCAGTCTGGGGTTCCTTCGTGCTCGGTTTGCGGTGAGCAGATCGGATTGGCATCA 79

Qy         21 AsnGlyGluLeuPheValAlaCysHisGluCysSerTyrProMetCysLysSerCysPhe 40
              |||||||||:::|||||||:|||||||:::| :::|:|||| :::|||||||||||||
Db         80 AATGGTGAAGTCTTTGTGGCTTGCCATGACTGCAACTACCCTCTCTGCAAGTCTTGCGTG 139

Qy         41 GluPheGluIleLysGluGlyArgLysValCysLeuArgCysGlySer 56
              |||:::|||:::|||||||:||||      ||||:::|:|||:|||:::
Db        140 GAGTATGAGCTGAAGGAAGGCCGCAACGCCTGCATGCGTTGTGGCACT 187

RESULT 4
BJ564652
LOCUS       BJ564652                 506 bp    mRNA    linear   EST 18-DEC-2002
DEFINITION  BJ564652 Ipomoea nil mixture of flower and flower bud Ipomoea nil
            cDNA clone jm5o14 5', mRNA sequence.
ACCESSION   BJ564652
VERSION     BJ564652.1  GI:27246472
KEYWORDS    EST.
SOURCE      Ipomoea nil (Japanese morning glory)
  ORGANISM  Ipomoea nil
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            asterids; lamiids; Solanales; Convolvulaceae; Ipomoea.
REFERENCE   1  (bases 1 to 506)
  AUTHORS   Hoshino,A., Seki,M., Shinozaki,T., Carninci,P., Kamiya,A., Shiraki,T.,
            Nitasaka,E., Shinozaki,K., Hayashizaki,Y., Kohara,Y. and Iida,S.
  TITLE     ESTs of Japanese morning glory
  JOURNAL   Unpublished
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .506
                     /organism="Ipomoea nil"
                     /mol_type="mRNA"
                     /cultivar="Tokyo-kokei standard"
                     /db_xref="taxon:35883"
                     /clone="jm5o14"
                     /tissue_type="mixture of flower and flower bud"
                     /clone_lib="Ipomoea nil mixture of flower and flower bud"
BASE COUNT      161 a    100 c    143 g    102 t
ORIGIN

Alignment Scores:
Pred. No.:              2.71e-20        Length:         506
Score:                  236.00          Matches:        36
Percent Similarity:     83.93%          Conservative:   11
Best Local Similarity:  64.29%          Mismatches:     9
Query Match:            72.84%          Indels:         0
DB:                     12              Gaps:           0

US-09-980-043-4 (1-56) x BJ564652 (1-506)

Qy          1 MetMetGluSerGlyAlaProIleCysHisThrCysGlyGluGlnValGlyHisAspAla 20
              ||||||:::||||||   |||   |||      ||||||||||||:::|||      :::
Db         20 ATGATGCAGTCTGGGGTTCCTTCGTGCTCGGTTTGCGGTGAGCAGATCGGATTGGCATCA 79

Qy         21 AsnGlyGluLeuPheValAlaCysHisGluCysSerTyrProMetCysLysSerCysPhe 40
              |||||||||:::|||||||||||||||:::|||:::||||||:::||||||||||||
Db         80 AATGGTGAAGTCTTTGTGGCTTGCCATGATTGCAACTACCCTCTCTGCAAGTCTTGCGTG 139

Qy         41 GluPheGluIleLysGluGlyArgLysValCysLeuArgCysGlySer 56
              |||:::|||:::||||||||||||      |||:::||||||||:::
Db        140 GAGTATGAGCTGAAGGAAGGCCGCAACGCCTGCATGCGTTGTGGCACT 187

RESULT 5
BJ554190
LOCUS       BJ554190                 511 bp    mRNA    linear   EST 18-DEC-2002
DEFINITION  BJ554190 Ipomoea nil mixture of flower and flower bud Ipomoea nil
            cDNA clone jm13f17 5', mRNA sequence.
ACCESSION   BJ554190
VERSION     BJ554190.1  GI:27236010
KEYWORDS    EST.
SOURCE      Ipomoea nil (Japanese morning glory)
  ORGANISM  Ipomoea nil
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            asterids; lamiids; Solanales; Convolvulaceae; Ipomoea.
REFERENCE   1  (bases 1 to 511)
  AUTHORS   Hoshino,A., Seki,M., Shin-i,T., Carninci,P., Kamiya,A., Shiraki,T.,
            Nitasaka,E., Shinozaki,K., Hayashizaki,Y., Kohara,Y. and Iida,S.
  TITLE     ESTs of Japanese morning glory
  JOURNAL   Unpublished
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .511
                     /organism="Ipomoea nil"
                     /mol_type="mRNA"
                     /cultivar="Tokyo-kokei standard"
                     /db_xref="taxon:35883"
                     /clone="jm13f17"
                     /tissue_type="mixture of flower and flower bud"
                     /clone_lib="Ipomoea nil mixture of flower and flower bud"
BASE COUNT      165 a    100 c    142 g    104 t
ORIGIN

Alignment Scores:
Pred. No.:              2.75e-20        Length:         511
Score:                  236.00          Matches:        36
Percent Similarity:     83.93%          Conservative:   11
Best Local Similarity:  64.29%          Mismatches:     9
Query Match:            72.84%          Indels:         0
DB:                     12              Gaps:           0

US-09-980-043-4 (1-56) x BJ554190 (1-511)

Qy          1 MetMetGluSerGlyAlaProIleCysHisThrCysGlyGluGlnValGlyHisAspAla 20
              ||||||:::||||||   |||   |||      ||||||||||||:::|||      :::
Db         20 ATGATGCAGTCTGGGGTTCCTTCGTGCTCGGTTTGCGGTGAGCAGATCGGATTGGCATCA 79

Qy         21 AsnGlyGluLeuPheValAlaCysHisGluCysSerTyrProMetCysLysSerCysPhe 40
              |||||||||:::|||||||||||||||:::|||:::||||||:::||||||||||||
Db         80 AATGGTGAAGTCTTTGTGGCTTGCCATGACTGCAACTACCCTCTCTGCAAGTCTTGCGTG 139

Qy         41 GluPheGluIleLysGluGlyArgLysValCysLeuArgCysGlySer 56
              |||:::|||:::||||||||||||      |||:::||||||||:::
Db        140 GAGTATGAGCTGAAGGAAGGCCGCAACGCCTGCATGCGTTGTGGCACT 187
```

```
ESULT 6
G135016
OCUS        BG135016                 378 bp    mRNA    linear   EST 31-JAN-2001
EFINITION   EST467908 tomato crown gall Lycopersicon esculentum cDNA clone
            cTOE21E21 5' sequence, mRNA sequence.
CCESSION    BG135016
ERSION      BG135016.1  GI:12635204
EYWORDS     EST.
OURCE       Lycopersicon esculentum (tomato)
 ORGANISM   Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            asterids; lamiids; Solanales; Solanaceae; Solanum; Lycopersicon.
EFERENCE    1  (bases 1 to 378)
 AUTHORS    van der Hoeven,R., Sun,H., Cho,J., Utterback,T., Hansen,C., Ronning
            ,C. and Tanksley,S.
 TITLE      Generation of ESTs from tomato crown gall tissue
 JOURNAL    Unpublished
OMMENT      Contact: CUGI
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Email: http://www.genome.clemson.edu/orders/index.html.
EATURES             Location/Qualifiers
     source         1..378
                    /organism="Lycopersicon esculentum"
                    /mol_type="mRNA"
                    /cultivar="TA496"
                    /db_xref="taxon:4081"
                    /clone="cTOE21E21"
                    /tissue_type="crown gall"
                    /dev_stage="crown galls from full-grown plants (8 wks old
                    )"
                    /lab_host="SOLR"
                    /clone_lib="tomato crown gall"
                    /note="Vector: pBluescript SK(-); Site_1: EcoR1; Site_2:
                    Xho1; Four wk old greenhouse plants were stab inoculated
                    on stem with Agrobacterium tumefaciens C58 (Dr. T.J. Burr,
                    Cornell U.). Galls were allowed to develop for another 4
                    wks, when gall tissue was frozen in  liquid nitrogen."
ASE COUNT      114 a     60 c     96 g    108 t
RIGIN

lignment Scores:
red. No.:                2.83e-20        Length:          378
core:                    234.50          Matches:         39
ercent Similarity:       84.21%          Conservative:    9
est Local Similarity:    68.42%          Mismatches:      8
uery Match:              72.38%          Indels:          1
B:                       10              Gaps:            1

S-09-980-043-4 (1-56) x BG135016 (1-378)

y           1 MetMetGluSerGlyAlaProIleCysHisThrCysGlyGluGlnValGly---HisAsp 19
              ||||||||||||||||   |||||||||:::   ||||||||||||||||||      :::
b          43 ATGATGGAGTCAGGGGTTCCAATATGCAACATTTGCGGCGAACAAGTGGGGTTAGCTAAT 102

y          20 AlaAsnGlyGluLeuPheValAlaCysHisGluCysSerTyrProMetCysLysSerCys 39
                 ||||||||||:::|||||||||||||||||||||:::|||||:::|||||:::|||
b         103 ACTAATGGTGAAGTCTTTGTGGCTTGTCATGAATGTAACTATCCTGTTTGTAAGACTTGC 162

y          40 PheGluPheGluIleLysGluGlyArgLysValCysLeuArgCysGlySer 56
                 ::::::||||||||||||||||||      ||||||||||||   :::
b         163 ATAGATTATGAGATCAAAGAAGGACGCAACGCATGTCTACGGTGTGCAACT 213

ESULT 7
G453374
OCUS        BG453374                 661 bp    mRNA    linear   EST 16-MAR-2001
EFINITION   NF091C07LF1F1051 Developing leaf Medicago truncatula cDNA clone
            NF091C07LF 5', mRNA sequence.
CCESSION    BG453374
ERSION      BG453374.1  GI:13372168
KEYWORDS    EST.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 661)
  AUTHORS   Torres-Jerez,I., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell,C.J.,
            Flores,H.E., Irwin,J.T., Weller,J.W. and May,G.D.
  TITLE     Expressed Sequence Tags from the Samuel Roberts Noble Foundation
            Medicago truncatula leaf library
  JOURNAL   Unpublished
COMMENT     Contact: May GD
            Plant Biology Division
            The Samuel Roberts Noble Foundation
            2510 Sam Noble Parkway, Ardmore, OK 73402, USA
            Tel: 580 221 7361
            Fax: 580 221 7380
            Email: gdmaysnobl@yahoo.com
            Insert Length: 611  Std Error: 0.00
            Plate: 091  row: P  column: 11
            Seq primer: TCACGACGCGAAACAGCTATGA
FEATURES             Location/Qualifiers
     source          1..661
                     /organism="Medicago truncatula"
                     /mol_type="mRNA"
                     /db_xref="taxon:3880"
                     /clone="NF091C07LF"
                     /tissue_type="leaf"
                     /dev_stage="Pooled developmental"
                     /clone_lib="Developing leaf"
                     /note="Vector: Lambda Zap; Contains a mixture of very
                     young, developing, mature and senescing leaves."
BASE COUNT      213 a    114 c    138 g    195 t      2 others
ORIGIN

Alignment Scores:
Pred. No.:                1.72e-19        Length:          661
Score:                    231.00          Matches:         37
Percent Similarity:       82.14%          Conservative:    9
Best Local Similarity:    66.07%          Mismatches:      10
Query Match:              71.30%          Indels:          0
DB:                       10              Gaps:            0

US-09-980-043-4 (1-56) x BG453374 (1-661)

Qy          1 MetMetGluSerGlyAlaProIleCysHisThrCysGlyGluGlnValGlyHisAspAla 20
              ||||||||||||||||          ||||:::::|||||||||||||||||    |||
Db        298 ATGATGGAATCTGGAGTTAGCTTTTGTAATTCTTGTGGTGAACAAGTCGGAGTTGATGGT 357

Qy         21 AsnGlyGluLeuPheValAlaCysHisGluCysSerTyrProMetCysLysSerCysPhe 40
              |||||'||:::|||||||||||||||||||||'   :::|||:::|||||:::|||
Db        358 AATGGAGAGGTGTTTGTGGCTTGTCATGAATGTTATTTCCCAATTTGTAAGGCTTGTGTT 417

Qy         41 GluPheGluIleLysGluGlyArgLysValCysLeuArgCysGlySer 56
              :::::::|      || |||: |    ||| |.||||||||||:::
Db        418 GATTATGAAACCAGTGAGGGACGTAGTGTTTGCTTGAGATGTGGCAAT 465

RESULT 8
BE326104
LOCUS       BE326104                 515 bp    mRNA    linear   EST 21-DEC 2000
DEFINITION  NF086F04ST1F1032 Developing stem Medicago truncatula cDNA clone
            NF086F04ST 5', mRNA sequence.
ACCESSION   BE326104
VERSION     BE326104.2  GI:11958873
KEYWORDS    EST.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
```

```
REFERENCE   1  (bases 1 to 515)
  AUTHORS   He,X.-Z., Shadle,G., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell
            ,C.J., Flores,H.R., Inman,J.T., Weller,J.W., May,G.D. and Dixon
            ,R.A.
  TITLE     Expressed Sequence Tags from the Samuel Roberts Noble Foundation
            Medicago truncatula stem library
  JOURNAL   Unpublished
COMMENT     On Jul 14, 2000 this sequence version replaced gi:9199881.
            Contact: Dixon RA
            Plant Biology Division
            The Samuel Roberts Noble Foundation
            2510 Sam Noble Parkway, Ardmore, OK 73402, USA
            Tel: 580 221 7302
            Fax: 580 221 7380
            Email: radixon@noble.org
            Medicago Genome Initiative accession: MGI:S:26688
            Insert Length: 674   Std Error: 0.00
            Plate: 086  row: F  column: 04
            Seq primer: TCACACAGGAAACAGCTATGAC.
FEATURES             Location/Qualifiers
     source          1. .515
                     /organism="Medicago truncatula"
                     /mol_type="mRNA"
                     /db_xref="taxon:3880"
                     /clone="NF086P04ST"
                     /tissue_type="stem"
                     /dev_stage="Pooled developmental"
                     /clone_lib="Developing stem"
                     /note="Vector: Lambda Zap; Contains a mixture of
                     internodal stem segments"
BASE COUNT      143 a     89 c    127 g    156 t
ORIGIN

Alignment Scores:
Pred. No.:                1.81e-13        Length:         515
Score:                    183.00          Matches:        27
Percent Similarity:       74.07%          Conservative:   14
Best Local Similarity:    50.00%          Mismatches:     14
Query Match:              56.48%          Indels:         0
DB:                       10              Gaps:           0

US-09-980-043-4 (1-56) x BE326104 (1-515)

Qy          3 GluSerGlyAlaProIleCysHisThrCysGlyGluGlnValGlyHisAspAlaAsnGly 22
              :::|||   :::   ||||||           |||||||||||:|:::      ||||||
Db        131 AAATCTTCATCGAAGATATGTCGA?TTT???G?GATGAGATT?GATATAAGGAAAATGGA 190

Qy         23 GluLeuPheValAlaCysHisGluCysSerTyrProMetCysLysSerCysPheGluPhe 42
              |:|:||||:|||||:||||  ||:|:::::|:|:  |:|:  |:::||:::
Db        191 GAGCTTTTTGTGGCTTGTCATA?GTGT??ATT?CTTGTTT?AAGCCTTGTTATGAGTAT 250

Qy         43 GluIleLysGluGlyArgLysValCysLeuArgCysGlySer 56
              |||      ||||.|  :::  :|  :::|||  ||
Db        251 GAGAGGAGTGAAGGGAATCAATGTTGCCCTCAATGCAACAGT 292

RESULT 9
CA919485
LOCUS       CA919485                 558 bp    mRNA    linear   EST 27-DEC-2002
DEFINITION  EST637203 MTUS Medicago truncatula cDNA clone MTUS-14C9, mRNA
            sequence.
ACCESSION   CA919485
VERSION     CA919485.1  GI:27406415
KEYWORDS    EST.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 558)
  AUTHORS   VandenBosch,K., Endre,G., Silverstein,K., Town,C.D., Van Aken,S.
            Utterback,T., Cheung,F. and Fraser,C.M.
  TITLE     The Medicago truncatula EST unigene set: cDNA clones selected and
            re-arrayed from various libraries
  JOURNAL   Unpublished
COMMENT     Contact: VandenBosch K
            Department of Plant Biology
            University of Minnesota
            220 BioSci Center, 1445 Gortner Ave, St. Paul, MN 55108, USA
            Tel: 612 624 2755
            Fax: 612 625 1738
            Email: kvandenb@cbs.umn.edu
            TIGR sequence name: MTUAN33TV
            More information is available at: www.medicago.org
            Seq primer: (gtA AtA CgA CtC ACt AtA ggg C).
FEATURES             Location/Qualifiers
     source          1. .558
                     /organism="Medicago truncatula"
                     /mol_type="mRNA"
                     /cultivar="A17"
                     /db_xref="taxon:3880"
                     /clone="MTUS-14C9"
                     /tissue_type="mixed tissues"
                     /dev_stage="various stages"
                     /lab_host="XLOLR"
                     /clone_lib="MTUS"
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA
                     was directionally ligated into the Unizap XR vector from
                     Stratagene and packaged using Gigapack III Gold packaging
                     extracts. Plasmids containing cDNA inserts were excised
                     from the recombinant lambda-Zap phage using Ex-assist
                     helper phage and propagated in XLOLR cells."
BASE COUNT      181 a     80 c    148 g    149 t
ORIGIN

Alignment Scores:
Pred. No.:                2.02e-13        Length:         558
Score:                    183.00          Matches:        27
Percent Similarity:       74.07%          Conservative:   13
Best Local Similarity:    50.00%          Mismatches:     14
Query Match:              56.48%          Indels:         0
DB:                       14              Gaps:           0

US-09-980-043-4 (1-56) x CA919485 (1-558)

Qy          3 GluSerGlyAlaProIleCysHisThrCysGlyGluGlnValGlyHisAspAlaAsnGly 22
              :::|||   :::   ||||||      ||||||:::::::::|||:::      ||||||
Db         69 AAATCTTCATCGAAGATATGTCGAGTTTGTGGCGATGAGATTGGATATAAGGAAAATGGA 128

Qy         23 GluLeuPheValAlaCysHisGluCysSerTyrProMetCysLysSerCysPheGluPhe 42
              ||||||||||||||||||||  |||:::::|||:::||||||   |||:::|||:::
Db        129 GAGCTTTTTGTGGCTTGTCATGTGTGTGCATTTCCTGTTTGTAAGCCTTGTTATGAGTAT 188

Qy         43 GluIleLysGluGlyArgLysValCysLeuArgCysGlySer 56
              |||      ||||||   :::   |||   :::|||   |||
Db        189 GAGAGGAGTGAAGGGAATCAATGTTGCCCTCAATGCAACAGT 230

RESULT 10
BE203672
LOCUS       BE203672                 500 bp    mRNA    linear   EST 05-SEP-2000
DEFINITION  EST396348 KV0 Medicago truncatula cDNA clone pKV0-11K17, mRNA
            sequence.
ACCESSION   BE203672
VERSION     BE203672.1  GI:8746949
KEYWORDS    EST.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 500)
  AUTHORS   VandenBosch,K., Endre,G., Hur,J., Moore,J., Beremand,P., Ellis,L.,
            Town,C.D., Bowman,C.L., Craven,M.B., Hansen,T.S., Holt,I.E. and
            Fraser,C.M.
```

```
 TITLE     ESTs from uninoculated seedling roots of Medicago truncatula
 JOURNAL   Unpublished
OMMENT     Contact: VandenBosch K
           Department of Plant Biology
           University of Minnesota
           220 BioSci Center, 1445 Gortner Ave, St. Paul, MN 55108, USA
           Tel: 612 624 2755
           Fax: 612 625 1738
           Email: kvandenb@cbs.umn.edu
           Texas A&M University name:T263444e
           TIGR sequence name:MTGAE69TK
           More information is available at.
           http://chrysie.tamu.edu/medicago
           Seq primer: SKmod (CTA gAA CTA gtg gAT CC).
EATURES              Location/Qualifiers
     source          1..500
                     /organism="Medicago truncatula"
                     /mol_type="mRNA"
                     /cultivar="genotype A17"
                     /db_xref="taxon:3880"
                     /clone="pKV0-11K17"
                     /tissue_type="Seedling roots"
                     /dev_stage="Immediately prior to inoculation with
                     Sinorhizobium meliloti (0 hour)"
                     /lab_host="E.coli strain XLOLR"
                     /clone_lib="KV0"
                     /note="Vector: pBluescript SK -; Site_1: EcoRI; Site_2:
                     XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA
                     was directionally ligated into the Unizap XR vector from
                     Stratagene and packaged using Gigapack III Gold packaging
                     extracts. Plasmids containing cDNA inserts were excised
                     from the recombinant lambda-Zap phage using Ex-assist
                     helper phage and propagated in XLOLR cells."
ASE COUNT      149 a      88 c      127 g      136 t
RIGIN

lignment Scores:
red. No.:              1.85e-12        Length:         500
core:                  175.00          Matches:        26
ercent Similarity:     74.47%          Conservative:   9
est Local Similarity:  55.32%          Mismatches:     12
uery Match:            54.01%          Indels:         0
B:                     10              Gaps:           0

/S-09-980-043-4 (1-56) x BE203672 (1-500)

Qy          8 IleCysHisThrCysGlyGluGlnValGlyHisAspAlaAsnGlyGluLeuPheValAla 27
              :::||||||   |||||||||   ||||||             |||::::::|||||||||
Db        217 GTCTGTCATATATGTGGTGAGGATGTTGGAACAACACCCACCGGTGATGTGTTTGTGGCT 276

Qy         28 CysHisGluCysSerTyrProMetCysLysSerCysPheGluPheGluIleLysGluGly 47
              |||:::||||||   ||||||:::|||:::   |||:::|||:::|||   |||||||||
Db        277 TGTAATGAATGTGGTTACCCTGTTTGTCGAGATTGTTATGAATATGAGCGAAAAGAAGGC 336

Qy         48 ArgLysValCysLeuArgCys 54
              |||   |||   :::|||
Db        337 AATAAATCTTGTCCTCAATGC 357

ESULT 11
W585621
OCUS       AW585621                533 bp    mRNA    linear   EST 07-SEP-2000
EFINITION  EST317244 MHAM Medicago truncatula/Glomus versiforme mixed EST
           library cDNA clone pMHAM-32N6, mRNA sequence.
CCESSION   AW585621
ERSION     AW585621.1  GI:7265135
EYWORDS    EST.
OURCE      Medicago truncatula/Glomus versiforme mixed EST library
  ORGANISM Medicago truncatula/Glomus versiforme mixed EST library
           Eukaryota; mixed EST libraries.
EFERENCE   1  (bases 1 to 533)
  AUTHORS  Harrison,M.J., Liu,J., Peng,H., Gonzales,M., Ellis,L., Town,C.D.,
           Bowman,C.L., Craven,M.B., Hansen,T.S., Holt,I.E. and Fraser,C.M.
  TITLE    ESTs from roots of Medicago truncatula after colonization with
           Glomus versiforme
  JOURNAL  Unpublished
COMMENT    Contact: Maria J. Harrison
           Plant Biology Division
           The Samuel Roberts Noble Foundation
           2510 Sam Noble Parkway, Ardmore, OK 73401, USA
           Tel: 580-223 5810
           Fax: 580 221-7380
           Email: mjharris@noble.org
           Noble EST name:N254549e
           TIGR sequence name:MTDAD75TK
           More information is available at.
           'http://chrysie.tamu.edu/medicago'
           Seq primer: SKmod (CTA gAA CTA gtg gAT CC).
FEATURES             Location/Qualifiers
     source          1..533
                     /organism="Medicago truncatula/Glomus versiforme mixed EST
                     library"
                     /mol_type="mRNA"
                     /cultivar="Medicago truncatula genotype A17"
                     /db_xref="taxon:3880"
                     /clone="pMHAM-32N6"
                     /tissue_type="roots co-cultured with Glomus versiforme"
                     /dev_stage="Roots harvested at 10, 17, 22, 31 and 38 days
                     post-inoculation with Glomus versiforme. The library was
                     made from a mixture of RNA from each of these stages."
                     /lab_host="E. coli strain XLOLR"
                     /clone_lib="MHAM"
                     /note="Vector: pBluescript SK -; Site_1: EcoRI; Site_2:
                     XhoI; cDNA was prepared from polyA+ enriched RNA from
                     roots harvested at 10, 17, 22, 31 and 38 days
                     post-inoculation with Glomus versiforme. The cDNA was
                     directionally ligated into the Unizap XR vector from
                     Stratagene and packaged using Gigapack III Gold packaging
                     extracts. Plasmids containing cDNA inserts were excised
                     from the recombinant lambda-Zap phage using Ex-assist
                     helper phage and propagated in XLOLR cells."
BASE COUNT     161 a      95 c      131 g      146 t
ORIGIN

Alignment Scores:
Pred. No.:             2.03e-12        Length:         533
Score:                 175.00          Matches:        26
Percent Similarity:    74.47%          Conservative:   9
Best Local Similarity: 55.32%          Mismatches:     12
Query Match:           54.01%          Indels:         0
DB:                    9               Gaps:           0

US-09-980-043-4 (1-56) x AW585621 (1-533)

Qy          8 IleCysHisThrCysGlyGluGlnValGlyHisAspAlaAsnGlyGluLeuPheValAla 27
              :::||||||   |||||||||   ||||||             |||::::::|||||||||
Db        229 GTCTGTCATATATGTGGTGAGGATGTTGGAACAACACCCACCGGTGATGTGTTTGTGGCT 288

Qy         28 CysHisGluCysSerTyrProMetCysLysSerCysPheGluPheGluIleLysGluGly 47
              |||:::||||||   ||||||:::|||:::   |||:::|||:::|||   |||||||||
Db        289 TGTAATGAATGTGGTTACCCTGTTTGTCGAGATTGTTATGAATATGAGCGAAAAGAAGGC 348

Qy         48 ArgLysValCysLeuArgCys 54
              |||   |||   :::|||
Db        349 AATAAATCTTGTCCTCAATGC 369

RESULT 12
BQ141412
LOCUS      BQ141412                537 bp    mRNA    linear   EST 26-APR 2002
DEFINITION NF019C12PH1F1098 Phoma-infected Medicago truncatula cDNA clone
           NF019C12PH 5', mRNA sequence.
ACCESSION  BQ141412
VERSION    BQ141412.1  GI:20277538
KEYWORDS   EST.
SOURCE     Medicago truncatula (barrel medic)
  ORGANISM Medicago truncatula
           Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
```

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids ; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.
REFERENCE 1 (bases 1 to 537)
AUTHORS Watson,B.S., Shin,H.-S., Lopez-Meyer,M., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell,C.J., Inman,J.T., Waugh,M.E., Sullivan,J.P., May,G.D. and Paiva,N.L.
TITLE Expressed Sequence Tags from the Samuel Roberts Noble Foundation Medicago truncatula Phoma-infected library
JOURNAL Unpublished
COMMENT Contact: Paiva NL
Plant Biology Division
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73402, USA
Tel: 580 221 7317
Fax: 580 221 7380
Email: nlpaiva@noble.org
Insert Length: 537 Std Error: 0.00
Plate: 019 row: C column: 12
Seq primer: TCACACAGGAAACAGCTATGAC.
FEATURES Location/Qualifiers
source 1..537
/organism="Medicago truncatula"
/mol_type="mRNA"
/db_xref="taxon:3880"
/clone="NF019C12PH"
/tissue_type="leaf"
/dev_stage="Pathogen-induced, young trifoliate"
/clone_lib="Phoma-infected"
/note="Vector: pBluescript SK(-); Young trifoliate leaves of Medicago truncatula were excised and dip-inoculated in a spore suspension of Phoma medicaginis, and incubated in humid dishes. Pools of leaves were harvested at 0, 15, and 30 minutes and 1, 2, 3, 6, 14, 24, 48, 72, and 96, hours, and used to prepare total RNA. cDNA was prepared from polyA+ enriched, pooled samples of equivalent amounts of total RNA from each sample. The cDNA was directionally ligated into the Uni Zap XR vector (Stratagene) and packaged using the Gigapack III Gold packaging extracts. Phagemids containing cDNA inserts were in vivo excised from the recombinant Uni-ZAP XR vector using ExAssist helper phage and the E. coli strain XL1-Blue MRF' (Stratagene). Excised plasmids were plated using SOLR cells."
BASE COUNT 164 a 89 c 136 g 146 t 2 others
ORIGIN

Alignment Scores:
Pred. No.: 2.05e-12 Length: 537
Score: 175.00 Matches: 26
Percent Similarity: 74.47% Conservative: 9
Best Local Similarity: 55.32% Mismatches: 12
Query Match: 54.01% Indels: 0
DB: 13 Gaps: 0

US-09-980-043-4 (1-56) x BQ141412 (1-537)

Qy 8 IleCysHisThrCysGlyGluGlnValGlyHisAspAlaAsnGlyGluLeuPheValAla 27
Db 260 GTCTGTCATATATGTGGTGAGGATGTTGGAACAACACCCACCGGTGATGTGTTTGTGGCT 319

Qy 28 CysHisGluCysSerTyrProMetCysLysSerCysPheGluPheGluIleLysGluGly 47
Db 320 TGTAATGAATGTGGTTACCCTGTTTGTCGAGATTGTTATGAATATGAGCGAAAAGAAGGC 379

Qy 48 ArgLysValCysLeuArgCys 54
Db 380 AATAAATCTTGTCCTCAATGC 400

RESULT 13
AJ548388/c
LOCUS AJ548388 584 bp mRNA linear EST 28-FEB-2003
DEFINITION AJ548388 MTAPHEU Medicago truncatula cDNA clone mtaehac110003a10, mRNA sequence.
ACCESSION AJ548388
VERSION AJ548388.1 GI:28612102
KEYWORDS EST.
SOURCE Medicago truncatula (barrel medic)
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids ; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.
REFERENCE 1 (bases 1 to 584)
AUTHORS Nyamsuren,O., Colditz,F., Rosendahl,S., Bekel,T., Meyer,F., Kuester ,H., Franken,P. and Krajinski,F.
TITLE First insights into the pathogenic interaction formed between the oomycete Aphanomyces euteiches and legumes using the model plant Medicago truncatula
JOURNAL Unpublished
COMMENT Contact: Krajinski F
LG Molekulargenetik
Herrenhaeuser Str. 2 D-30419 Hannover, Germany.
FEATURES Location/Qualifiers
source 1..584
/organism="Medicago truncatula"
/mol_type="mRNA"
/db_xref="taxon:3880"
/clone="mtaehac110003a10"
/tissue_type="root"
/dev_stage="6 days afer inoculation with Aphanomyces euteiches zoospores"
/clone_lib="MTAPHEU"
/note="Vector: pGEM-T; Site_1: EcoRI; genotype A17; Total RNA was isolated from roots harvested 6 days after inoculation with Aphanomyces euteiches zoospores. cDNA was prepared from total RNA using the SMART PCR cDNA system (Clontech). This cDNA was used as tester in Suppression Substractive Hybridization (SSH). The SSH-cDNA fragments were generated using the SSH-adaptor sequences ctaatacgactcactatagggctcgagcggccgcccgggcaggt and ctaatacgactcactatagggcagcgtggtcgcggccgaggt (Clontech) and ligated after Suppression Subtractive Hybridization into the pGEM-Teasy vector (Promega). Plasmids containing cDNA inserts were propagated in E. coli TOP 10F' cells (InVitrogen)."
BASE COUNT 147 a 143 c 108 g 186 t
ORIGIN

Alignment Scores:
Pred. No.: 2.3e-12 Length: 584
Score: 175.00 Matches: 26
Percent Similarity: 74.47% Conservative: 9
Best Local Similarity: 55.32% Mismatches: 12
Query Match: 54.01% Indels: 0
DB: 9 Gaps: 0

US-09-980-043-4 (1-56) x AJ548388 (1-584)

Qy 8 IleCysHisThrCysGlyGluGlnValGlyHisAspAlaAsnGlyGluLeuPheValAla 27
Db 573 GTCTGTCATATATGTGGTGAGGATGTTGGAACAACACCCACCGGTGATGTGTTTGTGGCT 514

Qy 28 CysHisGluCysSerTyrProMetCysLysSerCysPheGluPheGluIleLysGluGly 47
Db 513 TGTAATGAATGTGGTTACCCTGTTTGTCGAGATTGTTATGAATATGAGCGAAAAGAAGGC 454

Qy 48 ArgLysValCysLeuArgCys 54
Db 453 AATAAATCTTGTCCTCAATGC 433

RESULT 14
AW693843
LOCUS AW693843 639 bp mRNA linear EST 20-DEC-2000
DEFINITION NF069F10ST1F1090 Developing stem Medicago truncatula cDNA clone NF069F10ST 5', mRNA sequence.

```
ACCESSION   AW693843
VERSION     AW693843.2  GI:11932270
KEYWORDS    EST.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 639)
  AUTHORS   He,X.-Z., Shadle,G., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell
            ,C.J., Flores,H.R., Inman,J.T., Weller,J.W., May,G.D. and Dixon
            ,R.A.
  TITLE     Expressed Sequence Tags from the Samuel Roberts Noble Foundation
            Medicago truncatula stem library
  JOURNAL   Unpublished
COMMENT     On Apr 14, 2000 this sequence version replaced gi:7568580.
            Contact: Dixon RA
            Plant Biology Division
            The Samuel Roberts Noble Foundation
            2510 Sam Noble Parkway, Ardmore, OK 73402, USA
            Tel: 580 221 7302
            Fax: 580 221 7380
            Email: radixon@noble.org
            Insert Length: 670   Std Error: 0.00
            Plate: 069  row: F  column: 10
            Seq primer: TCACACAGGAAACAGCTATGAC.
FEATURES             Location/Qualifiers
     source          1..639
                     /organism="Medicago truncatula"
                     /mol_type="mRNA"
                     /db_xref="taxon:3880"
                     /clone="NF069F10ST"
                     /tissue_type="stem"
                     /dev_stage="Pooled developmental"
                     /clone_lib="Developing stem"
                     /note="Vector: Lambda Zap; Contains a mixture of
                     internodal stem segments"
BASE COUNT      191 a    120 c    153 g    175 t
ORIGIN

Alignment Scores:
Pred. No.:              2.6e-12       Length:         639
Score:                  175.00        Matches:        26
Percent Similarity:     74.47%        Conservative:   9
Best Local Similarity:  55.32%        Mismatches:     12
Query Match:            54.01%        Indels:         0
DB:                     9             Gaps:           0

US-09-980-043-4 (1-56) x AW693843 (1-639)

Qy          8 IleCysHisThrCysGlyGluGlnValGlyHisAspAlaAsnGlyGluLeuPheValAla 27
              :::||||||   |||||||||   ||||||           |||:::::|||||||||
Db        258 GTCTGTCATATATGTGGTGAGGATGTTGGAACAACACCCACCGGTGATGTGTTTGTGGCT 317

Qy         28 CysHisGluCysSerTyrProMetCysLysSerCysPheGluPheGluIleLysGluGly 47
              |||:::||||||   ||||||:::|||:::   |||:::|||:::|||   ||||||||
Db        318 TGTAATGAATGTGGTTACCCTGTTTGTCGAGATTGTTATGAATATGAGCGAAAAGAAGGC 377

Qy         48 ArgLysValCysLeuArgCys 54
              |||   |||   :::|||
Db        378 AATAAATCTTGTCCTCAATGC 398

RESULT 15
BI307877
LOCUS       BI307877                643 bp    mRNA    linear   EST 20-JUL-2001
DEFINITION  EST529287 GPOD Medicago truncatula cDNA clone pGPOD-1O3 5' end,
            mRNA sequence.
ACCESSION   BI307877
VERSION     BI307877.1  GI:14982204
KEYWORDS    EST.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 643)
  AUTHORS   Grusak,M.A., Samac,D.A., Town,C.D., Van Aken,S., Utterback,T., Cho
            ,J. and Fraser,C.M.
  TITLE     ESTs from developing reproductive tissues of Medicago truncatula
  JOURNAL   Unpublished
COMMENT     Contact: Michael A. Grusak
            USDA/ARS Children's Nutrition Research Center
            Baylor College of Medicine
            1100 Bates Street, Houston, TX 77030-2600, USA
            Tel: 713-798-7044
            Fax: 713-798-7028
            Email: mgrusak@bcm.tmc.edu
            B394732e
            TIGR Sequence [illegible]
            More information is available at www.medicago.org
            Seq primer: SKm 1 (CTA GAA [illegible] CC)
FEATURES             Location/Qualifiers
     source          1..643
                     /organism="Medicago truncatula"
                     /mol_type="mRNA"
                     /cultivar="A17"
                     /db_xref="taxon:3880"
                     /clone="pGPOD-1O3"
                     /tissue_type="immature pod walls"
                     /dev_stage="Immature pods, ranging in age from 15 to 30
                     days after pollination"
                     /clone_lib="GPOD"
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI; Immature pods, ranging in age from 15 to 30 days
                     after pollination, were collected from greenhouse-grown
                     plants. At harvest, seeds were removed from pods and
                     isolated pod walls were collected and immediately frozen
                     in liquid nitrogen. Pod walls were pooled for mRNA
                     extraction. cDNA was prepared from polyA+ enriched RNA.
                     The cDNA was directionally ligated into the Unizap XR
                     vector from Stratagene and packaged using Gigapack III
                     Gold packaging extracts. Plasmids containing cDNA inserts
                     were excised from the recombinant lambda Zap phage using
                     Ex-assist helper phage and propogated in XLOLR cells."
BASE COUNT      193 a    134 c    147 g    169 t
ORIGIN

Alignment Scores:
Pred. No.:              2.62e-12      Length:         643
Score:                  175.00        Matches:        26
Percent Similarity:     74.47%        Conservative:   9
Best Local Similarity:  55.32%        Mismatches:     12
Query Match:            54.01%        Indels:         0
DB:                     12            Gaps:           0

US-09-980-043-4 (1-56) x BI307877 (1-643)

Qy          8 IleCysHisThrCysGlyGluGlnValGlyHisAspAlaAsnGlyGluLeuPheValAla 27
              :::||||||   |||||||||   ||||||           |||:::::|||||||||
Db        127 GTCTGTCATATATGTGGTGAGGATGTTGGAACAACACCCACCGGTGATGTGTTTGTGGCT 186

Qy         28 CysHisGluCysSerTyrProMetCysLysSerCysPheGluPheGluIleLysGluGly 47
              |||:::||||||   ||||||:::|||:::   |||:::|||:::|||   ||||||||
Db        187 TGTAATGAATGTGGTTACCCTGTTTGTCGAGATTGTTATGAATATGAGCGAAAAGAAGGC 246

Qy         48 ArgLysValCysLeuArgCys 54
              |||   |||   :::|||
Db        247 AATAAATCTTGTCCTCAATGC 267


Search completed: October 31, 2003, 17:42:43
Job time : 340.263 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 30, 2003, 16:11:59 ; Search time 12264.5 Seconds
(without alignments)
11487.872 Million cell updates/sec

Title: US-09-980-043-5
Perfect score: 3444
Sequence: 1 gcggccgcggttaatcgccg.....gtgttgtagatagcggccgc 3444

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2888711 seqs, 20454813386 residues

Word size : 0

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	Query Match	Length	DB	ID	Description
1		3444	100.0	3444	8	AF062485	AF062485 Arabidops
2		2322	67.4	3255	6	AX507835	AX507835 Sequence
3		971	28.2	1492	8	AF375439	AF375439 Arabidops
4		827	24.0	1101	8	AY143957	AY143957 Arabidops
5		773	22.4	39645	8	AB025637	AB025637 Arabidops
6		50	1.5	82390	8	AB016893	AB016893 Arabidops
7		39	1.1	107848	8	AC007019	AC007019 Arabidops
8		38	1.1	1041	8	BT008832	BT008832 Arabidops
9		38	1.1	1490	8	AY136423	AY136423 Arabidops
10	c	38	1.1	103966	8	AC125473	AC125473 Medicago
11		38	1.1	113199	8	AC126014	AC126014 Medicago
12		34	1.0	1838	8	AF030052	AF030052 Oryza sat
13		32	0.9	3222	6	AX652952	AX652952 Sequence
14		32	0.9	3255	6	AX505864	AX505864 Sequence
15		32	0.9	3771	8	AF027173	AF027173 Arabidops
16		32	0.9	3828	6	AR267558	AR267558 Sequence
17		32	0.9	3828	6	AX030944	AX030944 Sequence
18		32	0.9	3828	6	BD022677	BD022677 Manipulat
19		32	0.9	5009	6	AR267556	AR267556 Sequence
20		32	0.9	5009	6	AX030941	AX030941 Sequence
21		32	0.9	5009	6	BD022675	BD022675 Manipulat
22		32	0.9	89035	8	ATT22F8	AL050351 Arabidops
23		32	0.9	154555	8	AC135958	AC135958 Oryza sat
24		32	0.9	198151	8	ATCHRIV91	AL161595 Arabidops
25		29	0.8	3186	8	GHU58283	U58283 Gossypium h
26		29	0.8	3328	6	AR164434	AR164434 Sequence
27		29	0.8	3676	8	AF200529	AF200529 Zea mays
28		29	0.8	3723	8	AF150630	AF150630 Gossypium
29		29	0.8	6886	8	AF413210	AF413210 Gossypium
30	c	29	0.8	130300	8	AP005824	AP005824 Oryza sat
31		29	0.8	146543	2	AP003945	AP003945 Oryza sat
32		29	0.8	162776	8	AP003237	AP003237 Oryza sat
33		28	0.8	105284	2	AC140546	AC140546 Medicago
34	c	27	0.8	50149	2	AC132494	AC132494 Oryza sat
35	c	27	0.8	84322	8	AP004509	AP004509 Lotus jap
36	c	27	0.8	142160	2	AC135426	AC135426 Oryza sat
37	c	27	0.8	157010	2	AC144738	AC144738 Oryza sat
38		26	0.8	2958	8	AF267742	AF267742 Arabidops
39		26	0.8	3725	8	AF200526	AF200526 Zea mays
40		26	0.8	3752	8	AF200525	AF200525 Zea mays
41		26	0.8	3795	8	AF200533	AF200533 Zea mays
42		26	0.8	3799	6	AX338680	AX338680 Sequence
43		26	0.8	3812	8	AF200532	AF200532 Zea mays
44		26	0.8	3847	8	AF304374	AF304374 Nicotiana
45		24	0.7	3081	8	AF088917	AF088917 Arabidops

ALIGNMENTS

RESULT 1
AF062485
LOCUS AF062485 3444 bp mRNA linear PLN 08-AUG-1998
DEFINITION Arabidopsis thaliana cellulose synthase mRNA, partial cds.
ACCESSION AF062485
VERSION AF062485.1 GI:3135610
KEYWORDS .
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 3444)
AUTHORS Wu,L., Joshi,C.P. and Chiang,V.L.
TITLE AraxCelA, a new member of the cellulose synthase gene family from

```
          Arabidopsis (Accession No. AF062485) (PGR98-114)
 JOURNAL  Plant Physiol. 117 (3), 1125 (1998)
EFERENCE  2  (bases 1 to 3444)
 AUTHORS  Joshi,C.
 TITLE    Direct Submission
 JOURNAL  Submitted (01-MAY-1998) Plant Biotechnology Research Center, School
          of Forestry and Wood Products, Michigan Technological University,
          Houghton, MI 49931, USA
CATURES             Location/Qualifiers
     source         1. .3444
                    /organism="Arabidopsis thaliana"
                    /mol_type="mRNA"
                    /cultivar="Columbia"
                    /db_xref="taxon:3702"
                    /clone_lib="ABRC Number CD4-16"
     CDS            <1. .3247
                    /codon_start=2
                    /product="cellulose synthase"
                    /protein_id="AAC29067.1"
                    /db_xref="GI:3135611"
                    /translation="RPRLIAGSHNRNEFVLINADENARIRSVQELSGQTCQICGDEIE
                    LTVDGEPFVACNECAFPVCRPCYEYERREGNQACPQCKTRFKRLKGSPRVEGDEEEDD
                    IDDLDNEFEYGNNGIGFDQVSEGMSISRRNSGFPQSDLDSAPPGSQIPLLTYGDEDVE
                    ISSDRHALIVPPSLGGHGNRVHPVSLSDPTVAAHRRLMVPQKDLAVYGYGSVAWKDRM
                    EEWKRKQNEKLQVVRHEGDPDFEDGDDADFPMMDEGRQPLSMKIPIKSSKINPYRMLI
                    VLRLVILGLFFHYRILHPVKDAYALWLISVICEIWFAVSWVLDQFPKWYPIERETYLD
                    RLSLRYEKEGKPSGLSPVDVFVSTVDPLKEPPLITANTVLSILAVDYPVDKVACYVSD
                    DGAAMLTFEALSETAEFARKWVPFCKKYCIEPRAPEWYFCHKMDYLKNKVHPAFVRER
                    RAMKRDYEEFKVKINALVATAQKVPEDGWTMQDGTPWPGNSVRDHPGMIQVFLGSDGV
                    RDVENNELPRLVYVSREKRPGFDHHKKAGAMNSLIRVSGVLSNAPYLLNVDCDHYINN
                    SKALREAMCFMMDPQSGKKICYVQFPQRFDGIDRHDRYSNRNVVFFDINMKGLDGLQG
                    PIYVGTGCVFRRQALYGFDAPKKKKGPRKTCNCWPKWCLLCFGSRKNRKAKTVAADKK
                    KKNREASKQIHALENIEEGRGHKVLNVEQSTEAMQMKLQKKYGQSPVFVASARLENGG
                    MARNASPACLLKEAIQVISRGYEDKTEWGKEIGWIYGSVTEDILTGSKMHSHGWRHVY
                    CTPKLAAFKGSAPINLSDRLHQVLRWALGSVEIFLSRHCPIWYGYGGGLKWLERLSYI
                    NSVVYPWTSLPLIVYCSLPAICLLTGKFIVPEISNYASILFMALFSSIAITGILEMQW
                    GKVGIDDWWRNEQFWVIGGVSAHLFALFQGLLKVLAGVDTNFTVTSKAADDGEFSDLY
                    LFKWTSLLIPPMTLLIINVIGVIVGVSDAISNGYDSWGPLFGRLFFALWVIIHLYPFL
                    KGLLGKQDRMPTIIVVWSILLASILTLLWVRVNPFVAKGGPILEICGLDCL"
ASE COUNT      903 a    664 c    860 g    1017 t
RIGIN

 Query Match             100.0%;  Score 3444;  DB 8;  Length 3444;
 Best Local Similarity   100.0%;  Pred. No. 0;
 Matches 3444;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

y          1 GCGGCCGCGGTTAATCGCCGGTTCTCACAACAGGAATGAGTTTGTCCTCATTAATGCCGA 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
b          1 GCGGCCGCGGTTAATCGCCGGTTCTCACAACAGGAATGAGTTTGTCCTCATTAATGCCGA 60

y         61 TGAGAATGCCCGAATAAGATCAGTCCAAGAGCTGAGTGGACAGACATGTCAAATCTGCAG 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
b         61 TGAGAATGCCCGAATAAGATCAGTCCAAGAGCTGAGTGGACAGACATGTCAAATCTGCAG 120

y        121 AGATGAGATCGAATTGACTGTTGATGGAGAACCGTTTGTGGCATGTAACGAATGTGCATT 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
b        121 AGATGAGATCGAATTGACTGTTGATGGAGAACCGTTTGTGGCATGTAACGAATGTGCATT 180

y        181 CCCTGTGTGTAGACCTTGCTATGAGTACGAAAGACGAGAAGGCAATCAAGCTTGTCCACA 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
b        181 CCCTGTGTGTAGACCTTGCTATGAGTACGAAAGACGAGAAGGCAATCAAGCTTGTCCACA 240

y        241 GTGCAAAACCCGTTTCAAACGTCTTAAAGGAAGTCCAAGAGTTGAAGGTGATGAAGAGGA 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
b        241 GTGCAAAACCCGTTTCAAACGTCTTAAAGGAAGTCCAAGAGTTGAAGGTGATGAAGAGGA 300

y        301 AGATGACATTGATGATTTAGACAATGAGTTTGAGTATGGAAATAATGGGATTGGATTTGA 360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
b        301 AGATGACATTGATGATTTAGACAATGAGTTTGAGTATGGAAATAATGGGATTGGATTTGA 360

y        361 TCAGGTTTCTGAAGGTATGTCAATCTCTCGTCGCAACTCCGGTTTCCCACAATCTGATTT 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
b        361 TCAGGTTTCTGAAGGTATGTCAATCTCTCGTCGCAACTCCGGTTTCCCACAATCTGATTT 420

Qy       421 GGATTCAGTTCCAGTTTTCTCAGATTCATTGTTGACTTACGGCGACGAGGACGTTGA 480
            ||| |||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db       421 GGATTCAGTTCCAGTTTTCTCAGATTCATTGTTGACTTACGGCGACGAGGACGTTGA 480

Qy       481 GATTTCTTTTGATAGAGATGCTCTTATTGTTCGTTCTTCACTTGGTGGTCATGGCAATAG 540
            ||||||||||| |||| |||||||||||||||||||||||||||||||||||||||||||
Db       481 GATTTCTTTTGATAGAGATGCTCTTATTGTTCGTTCTTCACTTGGTGGTCATGGCAATAG 540

Qy       541 AGTTCATTTTGTTTTTTCTTCTGACTTGAGTCGTTGCTGCACATCGAAGGCTGATGGTACC 600
            ||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db       541 AGTTCATTTTGTTTTTTCTTCTGACTTGAGTCGTTGCTGCACATCGAAGGCTGATGGTACC 600

Qy       601 TCAGAAAGATCTTTTTGTTTATTTGTTATTGAAGTGTTGCTTGGAAAGATCGGATGGAGGA 660
            ||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db       601 TCAGAAAGATCTTTTTGTTTATTTGTTATTGAAGTGTTGCTTGGAAAGATCGGATGGAGGA 660

Qy       661 ATGGAAGAAAAGCAAAACTAAAGAAAAGCTTAGTCTTTAGGCATGAAGGAGATCTTGATTT 720
            ||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db       661 ATGGAAGAAAAGCAAAACTAAAGAAAAGCTTAGTCTTTAGGCATGAAGGAGATCTTGATTT 720

Qy       721 TGAAGATGGTGATGATGGTGATTTTCCAATGATGGATGAGGGAAGGCAGCCATTGTCTAT 780
            ||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db       721 TGAAGATGGTGATGATGGTGATTTTCCAATGATGGATGAGGGAAGGCAGCCATTGTCTAT 780

Qy       781 GAAGATAGCAATCAAATCGAGCAAGATAAATCCTTACCGGATGTTAATTGTGCTACGTCT 840
            ||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db       781 GAAGATAGCAATCAAATCGAGCAAGATAAATCCTTACCGGATGTTAATTGTGCTACGTCT 840

Qy       841 TGTGATTCTTGGTCTCTTCTTTCACTACCGTATTCTTCACCCCGTCAAAGATGCATATGC 900
            ||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db       841 TGTGATTCTTGGTCTCTTCTTTCACTACCGTATTCTTCACCCCGTCAAAGATGCATATGC 900

Qy       901 TTTGTGGGTTATTTCTGTTATATGTGAGATATGGTTTGCTGTTTCATGGGTTCTTGATCA 960
            ||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db       901 TTTGTGGGTTATTTCTGTTATATGTGAGATATGGTTTGCTGTTTCATGGGTTCTTGATCA 960

Qy       961 GTTCCCTAAATGGTACCCTATCGAGCGAGAAACGTACTTGGACCGACTCTCATTAAGATA 1020
            ||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db       961 GTTCCCTAAATGGTACCCTATCGAGCGAGAAACGTACTTGGACCGACTCTCATTAAGATA 1020

Qy      1021 TGAGAAAGAAGGGAAACCGTCGGGACTATCCCCTGTGGATGTATTTGTTAGTACAGTGGA 1080
            ||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db      1021 TGAGAAAGAAGGGAAACCGTCGGGACTATCCCCTGTGGATGTATTTGTTAGTACAGTGGA 1080

Qy      1081 TCCATTGAAAGAGCCTCCGCTTATTACTGCAAATACTGTCTTGTCTATTCTTGCTGTTGA 1140
            ||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db      1081 TCCATTGAAAGAGCCTCCGCTTATTACTGCAAATACTGTCTTGTCTATTCTTGCTGTTGA 1140

Qy      1141 TTATCCTGTCGATAAGGTTGCTTGTTACGTATCTGATGATGGTGCTGCTATGCTTACTTT 1200
            ||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db      1141 TTATCCTGTCGATAAGGTTGCTTGTTACGTATCTGATGATGGTGCTGCTATGCTTACTTT 1200

Qy      1201 CGAAGCTCTTTCTGAGACCGCTGAATTCGCAAGGAAATGGGTTCCTTTCTGCAAGAAATA 1260
            ||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db      1201 CGAAGCTCTTTCTGAGACCGCTGAATTCGCAAGGAAATGGGTTCCTTTCTGCAAGAAATA 1260

Qy      1261 TTGTATTGAGCCTCGTGCTCCCGAATGGTATTTCTGCCATAAAATGGACTACTTGAAGAA 1320
            ||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db      1261 TTGTATTGAGCCTCGTGCTCCCGAATGGTATTTCTGCCATAAAATGGACTACTTGAAGAA 1320

Qy      1321 TAAAGTTCATCCCGCATTTGTTAGGGAGCGGCGAGCCATGAAGAGAGATTATGAAGAATT 1380
            ||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db      1321 TAAAGTTCATCCCGCATTTGTTAGGGAGCGGCGAGCCATGAAGAGAGATTATGAAGAATT 1380

Qy      1381 CAAAGTAAAGATCAATGCTTTAGTAGCAACAGCACAGAAAGTGCCTGAGGATGGTTGGAC 1440
            ||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db      1381 CAAAGTAAAGATCAATGCTTTAGTAGCAACAGCACAGAAAGTGCCTGAGGATGGTTGGAC 1440

Qy      1441 TATGCAAGACGGTACACCTTGGCCCGGTAATAGTGTGCGAGATCATCCTGGCATGATTCA 1500
            ||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db      1441 TATGCAAGACGGTACACCTTGGCCCGGTAATAGTGTGCGAGATCATCCTGGCATGATTCA 1500
```

```
Qy       1501 GGTCTTCCTTGGAAGTGACGGTGTTCGTGATGTCGAAAACAACGAGTTGCCTCGATTAGT 1560
Db       1501 GGTCTTCCTTGGAAGTGACGGTGTTCGTGATGTCGAAAACAACGAGTTGCCTCGATTAGT 1560

Qy       1561 TTACGTTTCTCGTGAGAAGAGACCCGGATTTGATCACCATAAGAAGGCTGGAGCTATGAA 1620
Db       1561 TTACGTTTCTCGTGAGAAGAGACCCGGATTTGATCACCATAAGAAGGCTGGAGCTATGAA 1620

Qy       1621 TTCCCTGATACGAGTCTCTGGGGTTCTATCAAATGCTCCTTACCTTCTGAATGTCGATTG 1680
Db       1621 TTCCCTGATACGAGTCTCTGGGGTTCTATCAAATGCTCCTTACCTTCTGAATGTCGATTG 1680

Qy       1681 TGATCACTACATCAACAATAGCAAAGCTCTTAGAGAAGCAATGTGTTTCATGATGGATCC 1740
Db       1681 TGATCACTACATCAACAATAGCAAAGCTCTTAGAGAAGCAATGTGTTTCATGATGGATCC 1740

Qy       1741 TCAGTCAGGAAAGAAAATCTGTTATGTTCAGTTCCCTCAAAGGTTCGATGGGATTGATAG 1800
Db       1741 TCAGTCAGGAAAGAAAATCTGTTATGTTCAGTTCCCTCAAAGGTTCGATGGGATTGATAG 1800

Qy       1801 GCACGATCGATACTCAAATCGCAATGTTGTGTTCTTTGATATCAATATGAAAGGTTTGGA 1860
Db       1801 GCACGATCGATACTCAAATCGCAATGTTGTGTTCTTTGATATCAATATGAAAGGTTTGGA 1860

Qy       1861 TGGGCTACAAGGGCCTATATACGTCGGTACAGGTTGTGTTTTCAGGAGGCAAGCGCTTTA 1920
Db       1861 TGGGCTACAAGGGCCTATATACGTCGGTACAGGTTGTGTTTTCAGGAGGCAAGCGCTTTA 1920

Qy       1921 CGGATTTGATGCACCGAAGAAGAAGAAGGGCCCACGTAAGACATGCAATTGCTGGCCAAA 1980
Db       1921 CGGATTTGATGCACCGAAGAAGAAGAAGGGCCCACGTAAGACATGCAATTGCTGGCCAAA 1980

Qy       1981 ATGGTGTCTCCTATGTTTTGGTTCAAGAAAGAATCGTAAAGTAAAGACAGTGGCTGCGGA 2040
Db       1981 ATGGTGTCTCCTATGTTTTGGTTCAAGAAAGAATCGTAAAGTAAAGACAGTGGCTGCGGA 2040

Qy       2041 TAAGAAGAAGAAGAATAGGGAAGCGTCAAAGCAGATCCACGCATTAGAAAATATCGAAGA 2100
Db       2041 TAAGAAGAAGAAGAATAGGGAAGCGTCAAAGCAGATCCACGCATTAGAAAATATCGAAGA 2100

Qy       2101 GGGCCGCGGTCATAAAGTTCTTAACGTAGAACAGTCAACCGAGGCAATGCAAATGAAGTT 2160
Db       2101 GGGCCGCGGTCATAAAGTTCTTAACGTAGAACAGTCAACCGAGGCAATGCAAATGAAGTT 2160

Qy       2161 GCAGAAGAAATATGGGCAGTCTCCTGTATTTGTTGCATCTGTGCGTCTGGAGAATGGTGG 2220
Db       2161 GCAGAAGAAATATGGGCAGTCTCCTGTATTTGTTGCATCTGTGCGTCTGGAGAATGGTGG 2220

Qy       2221 GATGGCTAGAAACGCAAGCGTCTTTTTTTCTAAAGAATTATCAATTCATTAGTCG 2280
Db       2221 GATGGCTAGAAACGCAAGCGTCTTTTTTTCTAAAGAATTATCAATTCATTAGTCG 2280

Qy       2281 CGGATATGAAGATAAAACTGAATGGGGAAAAGAGATTGGGTTGATCTATGGTTCTGTTAC 2340
Db       2281 CGGATATGAAGATAAAACTGAATGGGGAAAAGAGATTGGGTTGATCTATGGTTCTGTTAC 2340

Qy       2341 CGAAGATATTCTTACGGGTTCTAAGATGCATTCTCATGGTTGGAGACATGTTTATTGTAC 2400
Db       2341 CGAAGATATTCTTACGGGTTCTAAGATGCATTCTCATGGTTGGAGACATGTTTATTGTAC 2400

Qy       2401 ACCAAAGTTAGCGGCTTTCAAAGGATCAGCTCCAATCAATCTTTCGGATCGTCTCCATCA 2460
Db       2401 ACCAAAGTTAGCGGCTTTCAAAGGATCAGCTCCAATCAATCTTTCGGATCGTCTCCATCA 2460

Qy       2461 AGTTCTTCGATGGGCGCTTGGGTCGGTTGAGATTTTCTTGATAGGCATTGTCCTATTTG 2520
Db       2461 AGTTCTTCGATGGGCGCTTGGGTCGGTTGAGATTTTCTTGATAGGCATTGTCCTATTTG 2520

Qy       2521 GTATGGTTATGGAGGTGGGTTGAAATGGCTTGAGCGGTTGTCTACATTAACTCTGTGGT 2580
Db       2521 GTATGGTTATGGAGGTGGGTTGAAATGGCTTGAGCGGTTGTCTACATTAACTCTGTGGT 2580

Qy       2581 TTACCCGTGGACCTCTCTACCGCTCATCGTTTACTGTTCTCTCCCTGCCATCTGTCTTCT 2640
Db       2581 TTACCCGTGGACCTCTCTACCGCTCATCGTTTACTGTTCTCTCCCTGCCATCTGTCTTCT 2640

Qy       2641 CACTGGAAAATTCATCGTTCCCGAGATTAGCAACTATGCGAGTATCCTCTTCATGGCGCT 2700
Db       2641 CACTGGAAAATTCATCGTTCCCGAGATTAGCAACTATGCGAGTATCCTCTTCATGGCGCT 2700

Qy       2701 CTTCTCGTCGATTGCAATAACGGGTATTCTCGAGATGCAATGGGGCAAAGTTGGGATCGA 2760
Db       2701 CTTCTCGTCGATTGCAATAACGGGTATTCTCGAGATGCAATGGGGCAAAGTTGGGATCGA 2760

Qy       2761 TGATTGGTGGAGAAACGAACAGTTTTGGGTCATTGGAGGTGTTTCTGCGCATCTGTTTGC 2820
Db       2761 TGATTGGTGGAGAAACGAACAGTTTTGGGTCATTGGAGGTGTTTCTGCGCATCTGTTTGC 2820

Qy       2821 TCTCTTCCAAGGTCTCCTCAAGGTTCTTGCTGGTGTCGACACTAACTTCACAGTCACATC 2880
Db       2821 TCTCTTCCAAGGTCTCCTCAAGGTTCTTGCTGGTGTCGACACTAACTTCACAGTCACATC 2880

Qy       2881 AAAAGCAGCTGATGATGGAGAGTTCTCTGACCTTTACCTCTTCAAATGGACTTCACTTCT 2940
Db       2881 AAAAGCAGCTGATGATGGAGAGTTCTCTGACCTTTACCTCTTCAAATGGACTTCACTTCT 2940

Qy       2941 CATCCCTCCAATGACTCTACTCATCATAAACGTCATTGGAGTCATAGTCGGAGTCTCTGA 3000
Db       2941 CATCCCTCCAATGACTCTACTCATCATAAACGTCATTGGAGTCATAGTCGGAGTCTCTGA 3000

Qy       3001 TGCCATCAGCAATGGATACGACTCGTGGGGACCGCTTTTCGGAAGACTGTTCTTTGCACT 3060
Db       3001 TGCCATCAGCAATGGATACGACTCGTGGGGACCGCTTTTCGGAAGACTGTTCTTTGCACT 3060

Qy       3061 TTGGGTCATCATTCATCTTTACCCGTTCCTTAAAGGTTTGCTTGGGAAACAAGATAGAAT 3120
Db       3061 TTGGGTCATCATTCATCTTTACCCGTTCCTTAAAGGTTTGCTTGGGAAACAAGATAGAAT 3120

Qy       3121 GCCAACCATTATTGTCGTCTGGTCCATCCTCCTGGCCTCGATTCTTACACTTCTTTGGGT 3180
Db       3121 GCCAACCATTATTGTCGTCTGGTCCATCCTCCTGGCCTCGATTCTTACACTTCTTTGGGT 3180

Qy       3181 CCGGGTTAATCCGTTTGTGGCGAAAGGCGGTCCTATTCTCGAGATCTGTGGTTTAGACTG 3240
Db       3181 CCGGGTTAATCCGTTTGTGGCGAAAGGCGGTCCTATTCTCGAGATCTGTGGTTTAGACTG 3240

Qy       3241 CTTGTGATTCGATTGACCGGTGGATGGGTTGGTGAAAAAGGTTTAATTCCCACGGATCAA 3300
Db       3241 CTTGTGATTCGATTGACCGGTGGATGGGTTGGTGAAAAAGGTTTAATTCCCACGGATCAA 3300

Qy       3301 AGAGAGGTAAGAGAGATATTGTTTTACCTCTAAAAGACTCCTTCATTGTGTTCATTAGAT 3360
Db       3301 AGAGAGGTAAGAGAGATATTGTTTTACCTCTAAAAGACTCCTTCATTGTGTTCATTAGAT 3360

Qy       3361 GATGAAAAATGAAAAGAAAAAGAAGATTTAATTTTGTTACGAGAATTGTTATTTTTGCAA 3420
Db       3361 GATGAAAAATGAAAAGAAAAAGAAGATTTAATTTTGTTACGAGAATTGTTATTTTTGCAA 3420

Qy       3421 GAATGTGTTGTAGATAGCGGCCGC 3444
Db       3421 GAATGTGTTGTAGATAGCGGCCGC 3444
```

```
RESULT 2
AX507835
LOCUS       AX507835               3255 bp    DNA     linear   PAT 27-SEP-2002
DEFINITION  Sequence 2530 from Patent WO0216655.
ACCESSION   AX507835
VERSION     AX507835.1  GI:23389072
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
```

```
 AUTHORS   Harper,J.F., Kreps,J., Wang,X. and Zhu,T.
 TITLE     Stress-regulated genes of plants, transgenic plants containing
           same, and methods of use
 JOURNAL   Patent: WO 0216655-A 2530 28-FEB-2002;
           The Scripps Research Institute (US) ; Syngenta Participations AG
           (CH)
EATURES             Location/Qualifiers
    source          1. .3255
                    /organism="Arabidopsis thaliana"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:3702"
ASE COUNT      841 a    641 c    813 g    960 t
RIGIN

 Query Match           67.4%;  Score 2322;  DB 6;  Length 3255;
 Best Local Similarity  99.4%;  Pred. No. 0;
 Matches 3222;  Conservative    0;  Mismatches   18;  Indels    0;  Gaps    0;

y          8 CGGTTAATCGCCGGTTCTCACAACAGGAATGAGTTTGTCCTCATTAATGCCGATGAGAAT 67
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
b         16 CGGTTAATCGCCGGTTCTCACAACAGGAATGAGTTTGTCCTCATTAATGCCGATGAGAAT 75

y         68 GCCCGAATAAGATCAGTCCAAGAGCTGAGTGGACAGACATGTCAAATCTGCAGAGATGAG 127
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
b         76 GCCCGAATAAGATCAGTCCAAGAGCTGAGTGGACAGACATGTCAAATCTGCAGAGATGAG 135

y        128 ATCGAATTGACTGTTGATGGAGAACCGTTTGTGGCATGTAACGAATGTGCATTCCCTGTG 187
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
b        136 ATCGAATTGACTGTTGATGGAGAACCGTTTGTGGCATGTAACGAATGTGCATTCCCTGTG 195

y        188 TGTAGACCTTGCTATGAGTACGAAAGACGAGAAGGCAATCAAGCTTGTCCACAGTGCAAA 247
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
b        196 TGTAGACCTTGCTATGAGTACGAAAGACGAGAAGGCAATCAAGCTTGTCCACAGTGCAAA 255

y        248 ACCCGTTTCAAACGTCTTAAAGGAAGTCCAAGAGTTGAAGGTGATGAAGAGGAAGATGAC 307
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
b        256 ACCCGTTTCAAACGTCTTAAAGGAAGTCCAAGAGTTGAAGGTGATGAAGAGGAAGATGAC 315

y        308 ATTGATGATTTAGACAATGAGTTTGAGTATGGAAATAATGGGATTGGATTTGATCAGGTT 367
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
b        316 ATTGATGATTTAGACAATGAGTTTGAGTATGGAAATAATGGGATTGGATTTGATCAGGTT 375

y        368 TCTGAAGGTATGTCAATCTCTCGTCGCAACTCCGGTTTCCCACAATCTGATTTGGATTCA 427
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
b        376 TCTGAAGGTATGTCAATCTCTCGTCGCAACTCCGGTTTCCCACAATCTGATTTGGATTCA 435

y        428 GCTCCACCTGGCTCTCAGATTCCATTGCTGACTTACGGCGACGAGGACGTTGAGATTTCT 487
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
b        436 GCTCCACCTGGCTCTCAGATTCCATTGCTGACTTACGGCGACGAGGACGTTGAGATTTCT 495

y        488 TCTGATAGACATGCTCTTATTGTTCCTCCTTCACTTGGTGGTCATGGCAATAGAGTTCAT 547
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
b        496 TCTGATAGACATGCTCTTATTGTTCCTCCTTCACTTGGTGGTCATGGCAATAGAGTTCAT 555

y        548 CCTGTTTCTCTTTCTGACCCGACCGTGGCTGCACATCGAAGGCTGATGGTACCTCAGAAA 607
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
b        556 CCTGTTTCTCTTTCTGACCCGACCGTGGCTGCACATCCAAGGCCTATGGTACCTCAGAAA 615

y        608 GATCTTGCGGTTTATGGTTATGGAAGTGTCGCTTGGAAAGATCGGATGGAGGAATGGAAG 667
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
b        616 GATCTTGCGGTTTATGGTTATGGAAGTGTCGCTTGGAAAGATCGGATGGAGGAATGGAAG 675

y        668 AGAAAGCAGAATGAGAAACTTCAGGTTGTTAGGCATGAAGGAGATCCTGATTTTGAAGAT 727
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
b        676 AGAAAGCAGAATGAGAAACTTCAGGTTGTTAGGCATGAAGGAGATCCTGATTTTGAAGAT 735

y        728 GGTGATGATGCTGATTTTCCAATGATGGATGAGGGAAGGCAGCCATTGTCTATGAAGATA 787
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
b        736 GGTGATGATGCTGATTTTCCAATGATGGATGAGGGAAGGCAGCCATTGTCTAGGAAGATA 795

y        788 CCAATCAAATCGAGCAAGATAAATCCTTACCGGATGTTAATTGTGCTACGTCTTGTGATT 847
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
b        796 CCAATCAAATCGAGCAAGATAAATCCTTACCGGATGTTAATTGTGCTACGTCTTGTGATT 855
```

```
Qy       848 CTTGGTCTCTTCTTTCACTACCGTATTCTTCACCCCGTCAAAGATGCATATGCTTTGTGG 907
            ||||||||||| ||||||||| |||| ||| |.||||||||||| |||||||||.|||
Db       856 CTTGGTCTCTTCTTTCACTACCGTATTCTTCACCCCGTCAAAGATGCATATGCTTTGTGG 915

Qy       908 CTTATTTCTGTTATATGTGAGATATGGTTTGCTGTTTCATGGGTTCTTGATCAGTTCCCT 967
            |||||||||||||||| ||| |||| |||||||||||||||||||||||||||||||||
Db       916 CTTATTTCTGTTATATGTGAGATATGGTTTGCTGTTTCATGGGTTCTTGATCAGTTCCCT 975

Qy       968 AAATGGTACCCTATCGAGCGAGAAACGTACTTGGACCGACTCTCATTAAGATATGAGAAA 1027
            |||||||||||||||| ||| |||||||||||||||||||||||||||||||||||||||
Db       976 AAATGGTACCCTATCGAGCGAGAAACGTACTTGGACCGACTCTCATTAAGATATGAGAAA 1035

Qy      1028 GAAGGGAAACCGTCGGGACTATCCCCTGTGGATGTATTTGTTAGTACAGTGGATCCATTG 1087
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1036 GAAGGGAAACCGTCGGGACTATCCCCTGTGGATGTATTTGTTAGTACAGTGGATCCATTG 1095

Qy      1088 AAAGAGCCTTCGCTACTACGAAATATGTTTGTCTATTCTTGCTGTTGATTATCCT 1147
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1096 AAAGAGCCTTCGCTACTACGAAATATGTTTGTCTATTCTTGCTGTTGATTATCCT 1155

Qy      1148 GTCGATAAGGTTGCTTGTTACGTATCTGATGATGGTGCTGCTATGCTTACTTTCGAAGCT 1207
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1156 GTCGATAAGGTTGCTTGTTACGTATCTGATGATGGTGCTGCTATGCTTACTTTCGAAGCT 1215

Qy      1208 CTTTCTGAGACCGCTGAATTCGCAAGGAAATGGGTTCCTTTCTGCAAGAAATATTGTATT 1267
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1216 CTTTCTGAGACCGCTGAATTCGCAAGGAAATGGGTTCCTTTCTGCAAGAAATATTGTATT 1275

Qy      1268 GAGCCTCGTGCTCCGGAATGGTATTTCTGCCATAAAATGGACTACTTGAAGAATAAAGTT 1327
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1276 GAGCCTCGTGCTCCGGAATGGTATTTCTGCCATAAAATGGACTACTTGAAGAATAAAGTT 1335

Qy      1328 CATCCCGCATTTGTTAGGGAGCGGCGAGCCATGAAGAGAGATTATGAAGAATTCAAAGTA 1387
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1336 CATCCCGCATTTGTTAGGGAGCGGCGAGCCATGAAGAGAGATTATGAAGAATTCAAAGTA 1395

Qy      1388 AAGATCAATGCTTTAGTAGCAACAGCACAGAAAGTGCCTGAGGATGGTTGGACTATGCAA 1447
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1396 AAGATCAATGCTTTAGTAGCAACAGCACAGAAAGTGCCTGAGGATGGTTGGACTATGCAA 1455

Qy      1448 GACGGTACACCTTGGCCCGGTAATAGTGTGCGAGATCATCCTGGCATGATTCAGGTCTTC 1507
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1456 GACGGTACACCTTGGCCCGGTAATAGTGTGCGAGATCATCCTGGCATGATTCAGGTCTTC 1515

Qy      1508 CTTGGAAGTGACGGTGTTCGTGATGTCGAAAACAACGAGTTGCCTCGATTAGTTTACGTT 1567
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1516 CTTGGAAGTGACGGTGTTCGTGATGTCGAAAACAACGAGTTGCCTCGATTAGTTTACGTT 1575

Qy      1568 TCTCGTGAGAAGAGACCCGGATTTGATCACCATAAGAAGGCTGGAGCTATGAATTCCCTG 1627
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1576 TCTCGTGAGAAGAGACCCGGATTTGATCACCATAAGAAGGCTGGAGCTATGAATTCCCTG 1635

Qy      1628 ATACGAGTCTCTGGGGTTCTATCAAATGCTCGTTACCTTCTGAATGTCGATTGTGATCAC 1687
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1636 ATACGAGTCTCTGGGGTTCTATCAAATGCTCGTTACCTTCTGAATGTCGATTGTGATCAC 1695

Qy      1688 TACATCAACAATAGCAAAGCTCTTAGAGAAGCAATGTGTTTCATGATGGATCCTCAGTCA 1747
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1696 TACATCAACAATAGCAAAGCTCTTAGAGAAGCAATGTGTTTCATGATGGATCCTCAGTCA 1755

Qy      1748 GGAAAGAAAATCTGTTATGTTCAGTTCCCTCAAAGGTTCGATGGGATTGATAGGCACGAT 1807
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1756 GGAAAGAAAATCTGTTATGTTCAGTTCCCTCAAAGGTTCGATGGGATTGATAGGCACGAT 1815

Qy      1808 CGATACTCAAATCGCAATGTTGTGTTCTTTGATATCAATATGAAAGGTTTGGATGGGCTA 1867
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1816 CGATACTCAAATCGCAATGTTGTGTTCTTTGATATCAATATGAAAGGTTTGGATGGGCTA 1875

Qy      1868 CAAGGGCCTATATACGTCGGTACAGGTTGTGTTTTCAGGAGGCAAGCGCTTTACGGATTT 1927
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1876 CAAGGGCCTATATACGTCGGTACAGGTTGTGTTTTCAGGAGGCAAGCGCTTTACGGATTT 1935
```

```
Qy       1928 GATGCACCGAAGAAGAAGAAGGGCCCACGTAAGACATGCAATTGCTGGCCAAAATGGTGT 1987
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1936 GATGCACCGAAGAAGAAGAAGGGCCCACGTAAGACATGCAATTGCTGGCCAAAATGGTGT 1995

Qy       1988 CTCCTATGTTTTGGTTCAAGAAAGAATCGTAAAGCAAAGACAGTGGCTGCGGATAAGAAG 2047
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1996 CTCCTATGTTTTGGTTCAAGAAAGAATCGTAAAGCAAAGACAGTGGCTGCGGATAAGAAG 2055

Qy       2048 AAGAAGAATAGGGAAGCGTCAAAGCAGATCCACGCATTAGAAAATATCGAAGAGGGCCGC 2107
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2056 AAGAAGAATAGGGAAGCGTCAAAGCAGATCCACGCATTAGAAAATATCGAAGAGGGCCGC 2115

Qy       2108 GGTCATAAAGTTCTTAACGTAGAACAGTCAACCGAGGCAATGCAAATGAAGTTGCAGAAG 2167
             |||| |||| | |||||||| |||| |||||||||| |||| |||||| |||||||||||
Db       2116 GTCACTAAAGGTTCTAACGTAGAACAGTCAACCGAGGCAATGCAAATGAAGTTGGAGAAG 2175

Qy       2168 AAATATGGGCAGTCTCCTGTATTTGTTGCATCTGCGCGTCTGGAGAATGGTGGGATGGCT 2227
             |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2176 AAATTTGGGCAGTCTCCTGTATTTGTTGCATCTGCGCGTATGGAGAATGGTGGGATGGCT 2235

Qy       2228 AGAAACGCAAGCCCGGCTTGTCTGCTTAAAGAAGCCATCCAAGTCATTAGTCGCGGATAT 2287
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2236 AGAAACGCAAGCCCGGCTTGTCTGCTTAAAGAAGCCATCCAAGTCATTAGTTGCGGATAT 2295

Qy       2288 GAAGATAAAACTGAATGGGGAAAAGAGATTGGGTGGATCTATGGTTCTGTTACCGAAGAT 2347
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2296 GAAGATAAAACTGAATGGGGAAAAGAGATTGGGTGGATCTATGGTTCTGTTACCGAAGAT 2355

Qy       2348 ATTCTTACGGGTTCTAAGATGCATTCTCATGGTTGGAGACATGTTTATTGTACACCAAAG 2407
             |||||||||||||| ||||||||||||||||||||||||| ||||||||||||||||||||
Db       2356 ATTCTTACGGGTTTTAAGATGCATTCTCATGGTTGGAGATCTGTTTATTGTACACCAAAG 2415

Qy       2408 TTAGCGGCTTTCAAAGGATCAGCTCCAATCAATCTTTCGGATCGTCTCCATCAAGTTCTT 2467
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2416 TTAGCGGCTTTCAAAGGATCAGCTCCAATCAATCTTTCGGATCGTCTCCATCAAGTTCTT 2475

Qy       2468 CGATGGGCGCTTGGGTCGGTTGAGATTTTCTTGAGTAGGCATTGTCCTATTTGGTATGGT 2527
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2476 CGATGGGCGCTTGGGTCGGTTGAGATTTTCTTGAGTAGGCATTGTCCTATTTGGTATGGT 2535

Qy       2528 TATGGAGGTGGGTTGAAATGGCTTGAGCGGTTGTCCTACATTAACTCTGTGGTTTACCCG 2587
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2536 TATGGAGGTGGGTTGAAATGGCTTGAGCGGTTGTCCTACATTAACTCTGTGGTTTACCCG 2595

Qy       2588 TGGACCTCTCTACCGCTCATCGTTTACTGTTCTCTCCCTGCCATCTGTCTTCTCACTGGA 2647
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2596 TGGACCTCTCTACCGCTCATCGTTTACTGTTCTCTCCCTGCCATCTGTCTTCTCACTGGA 2655

Qy       2648 AAATTCATCGTTCCCGAGATTAGTAACTATTGGAGTATGCTTCATGGCGCTCTTCTCG 2707
             |||||| |||||| ||| |||| |||| ||| |||| ||| |||| ||| ||
Db       2656 AAATTCATCGTTCCCGAGATTAGTAACTATTGGAGTATGCTTCATGGCGCTCTTCTCG 2715

Qy       2708 TCGATTGCAATAACGGGTATTCTCGAGATGCAATGGGGCAAAGTTGGGATCGATGATTGG 2767
             |||||||||||||||||||| |||| |||||||||| |||||||||||||||||||||||
Db       2716 TCGATTGCAATAACGGGTATTCTCGAGATGCAATGGGGCAAAGTTGGGATCGATGATTGG 2775

Qy       2768 TGGAGAAACGAACAGTTTTGGGTCATTGGAGGTGTTTCTGCTGCATCTGTTTGCTCTCTTC 2827
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2776 TGGAGAAACGAACAGTTTTGGGTCATTGGAGGTGTTTCTGCTGCATCTGTTTGCTCTCTTC 2835

Qy       2828 CAAGGTCTCCTCAAGGTTCTTGCTGGTGTCTGACACTAACTTCACAGTCACATCAAAAGCA 2887
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2836 CAAGGTCTCCTCAAGGTTCTTGCTGGTGTCTGACACTAACTTCACAGTCACATCAAAAGCA 2895

Qy       2888 GCTGATGATGGAGAGTTCTCTGACCTTTACCTCTTCAAATGGACTTCACTTCTCATCCCT 2947
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2896 GCTGATGATGGAGAGTTCTCTGACCTTTACCTCTTCAAATGGACTTCACTTCTCATCCCT 2955

Qy       2948 CCAATGACTCTACTCATCATAAACGTCATTGGAGTCATAGTCGGAGTCTCTGATGCCATC 3007
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2956 CCAATGACTCTACTCATCATAAACGTCATTGGAGTCATAGTCGGAGTCTCTGATGCCATC 3015

Qy       3008 AGCAATGGATACGACTCGTGGGGACCGCTTTTCGGAAGACTGTTCTTTGCACTTTGGGTC 3067
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3016 AGCAATGGATACGACTCGTGGGGACCGCTTTTCGGAAGACTGTTCTTTGCACTTTGGGTC 3075

Qy       3068 ATCATTCATCTTTACCCGTTCCTTAAAGGTTTGCTTGGGAAACAAGATAGAATGCCAACC 3127
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3076 ATCATTCATCTTTACCCGTTCCTTAAAGGTTTGCTTGGGAAACAAGATAGAATGCCAACC 3135

Qy       3128 ATTATTGTCGTCTGGTCCATCCTCCTGGCCTCGATTCTTACACTTCTTTGGGTCCGGGTT 3187
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3136 ATTATTGTCGTCTGGTCCATCCTCCTGGCCTCGATTCTTACACTTCTTTGGGTCCGGGTT 3195

Qy       3188 AATCCGTTTGTGGCGAAAGGCGGTCCTATTCTCGAGATCTGTGGTTTAGACTGCTTGTGA 3247
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3196 AATCCGTTTGTGGCGAAAGGCGGTCCTATTCTCGAGATCTGTGGTTTAGACTGCTTGTGA 3255
```

RESULT 3
AF375439

```
LOCUS       AF375439               1492 bp    mRNA    linear   PLN 07-FEB-2002
DEFINITION  Arabidopsis thaliana AT5g64740/MVP7_7 mRNA, complete cds.
ACCESSION   AF375439
VERSION     AF375439.1  GI:14030696
KEYWORDS    FLI_CDNA.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 1492)
  AUTHORS   Cheuk,R., Chen,H., Kim,C.J., Koesema,E., Meyers,M.C., Shinn,P.,
            Banh,J., Bowser,L., Carninci,P., Chung,M.K., Goldsmith,A.D.,
            Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G.,
            Kawai,J., Lam,B., Lee,J.M., Lin,J., Liu,S.X., Miranda,M.,
            Narusaka,M., Nguyen,M., Palm,C.J., Pham,P.K., Quach,H.L.,
            Sakano,H., Sakurai,T., Satou,M., Seki,M., Southwick,A., Tang,C.C.,
            Toriumi,M., Yamada,K., Yu,G., Shinozaki,K., Davis,R.W.,
            Theologis,A. and Ecker,J.R.
  TITLE     Arabidopsis cDNA clones
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1492)
  AUTHORS   Cheuk,R., Chen,H., Kim,C.J., Koesema,E., Meyers,M.C., Shinn,P.,
            Banh,J., Bowser,L., Carninci,P., Chung,M.K., Goldsmith,A.D.,
            Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G.,
            Kawai,J., Lam,B., Lee,J.M., Lin,J., Liu,S.X., Miranda,M.,
            Narusaka,M., Nguyen,M., Palm,C.J., Pham,P.K., Quach,H.L.,
            Sakano,H., Sakurai,T., Satou,M., Seki,M., Southwick,A., Tang,C.C.,
            Toriumi,M., Yamada,K., Yu,G., Shinozaki,K., Davis,R.W.,
            Theologis,A. and Ecker,J.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAY-2001) Salk Institute Genomic Analysis Laboratory
            (SIGnAL), Plant Biology Laboratory, The Salk Institute for
            Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037,
            USA
COMMENT     RIKEN Genomic Sciences Center (GSC) members carried out the
            collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN
            Arabidopsis Full-Length cDNA') :  Seki,M., Narusaka,M., Ishida,J.,
            Satou,M., Kamiya,A., Sakurai,T., Carninci,P.,  Kawai,J.,
            Hayashizaki,Y. and Shinozaki,K.

            The Salk, Stanford, PGEC (SSP) Consortium members carried out the
            sequencing and annotation of the RAFL cDNAs: Shinn,P., Chen,H.,
            Cheuk,R., Kim,C.J., Koesema,E., Meyers,M.C., Tracy,S.E., Banh,J.
            Bowser,L., Chung,M.K., Goldsmith,A.D., Jones,T., Karlin-Neumann,G.,
            Lam,B., Lee,J.M., Lin,J., Liu,S.X., Miranda,M., Nguyen,M.,
            Palm,C.J., Pham,P.K., Quach,H.L., Sakano,H., Southwick,A.,
            Tang,C.C., Toriumi,M., Yamada,K., Yu,G., Davis,R.W., Theologis,A.,
            and Ecker,J.R.

            Shinn,P. (SSP/Salk) and Seki,M. (RIKEN GSC) contributed equally to
            this work.  Shinozaki,K. (RIKEN GSC) and Ecker,J.R. (SSP/Salk)
            contributed equally to this work as PIs.
FEATURES             Location/Qualifiers
     source          1. .1492
```

```
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
                     /chromosome="5"
                     /clone="RAFL05-02-P19(R12458)"
                     /note="ecotype: Columbia"
     CDS             247. .1347
                     /note="cellulose synthase catalytic subunit"
                     /codon_start=1
                     /product="AT5g64740/MVP7_7"
                     /protein_id="AAK53023.1"
                     /db_xref="GI:14030697"
                     /translation="MQMKLEKKFGQSPVFVASARMENGGMARNASPACLLKEAIQVIS
                     CGYEDKTEWGKEIGWIYGSVTEDILTGFKMHSHGWRSVYCTPKLAAFEGSAPINLSDR
                     LHQVLRWALGSVEIFLSRHCPIWYGYGGGLKWLERLSYINSVVYPWTSLPLIVYCSLP
                     AICLLTGKFIVPEISNYASILFMALFSSIAITGILEMQWGKVGIDDWWRNEQFWVIGG
                     VSAHLFALFQGLLKVLAGVDTNFTVTSKAADDGEFSDLYLFKWTSLLIPPMTLLIINV
                     IGVIVGVSDAISNGYDSWGPLFGRLFFALWVIIHLYPFLKGLLGKQDRMPTIIVVWSI
                     LLASILTLLWVRVNPFVAKGGPILEICGLDCL"
ASE COUNT      383 a    298 c    373 g    438 t
IGIN

Query Match             28.2%;  Score 971;  DB 8;  Length 1492;
Best Local Similarity   100.0%;  Pred. No. 0;
Matches  971;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Y       2421 AAGGATCAGCTCCAATCAATCTTTCGGATCGTCTCCATCAAGTTCTTCGATGGGCGCTTG 2480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       521 AAGGATCAGCTCCAATCAATCTTTCGGATCGTCTCCATCAAGTTCTTCGATGGGCGCTTG 580

Y       2481 GGTCGGTTGAGATTTTCTTGAGTAGGCATTGTCCTATTTGGTATGGTTATGGAGGTGGGT 2540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       581 GGTCGGTTGAGATTTTCTTGAGTAGGCATTGTCCTATTTGGTATGGTTATGGAGGTGGGT 640

Y       2541 TGAAATGGCTTGAGCGGTTGTCCTACATTAACTCTGTGGTTTACCCGTGGACCTCTCTAC 2600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       641 TGAAATGGCTTGAGCGGTTGTCCTACATTAACTCTGTGGTTTACCCGTGGACCTCTCTAC 700

Y       2601 CGCTCATCGTTTACTGTTCTCTCCCTGCCATCTGTCTTCTCACTGGAAAATTCATCGTTC 2660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       701 CGCTCATCGTTTACTGTTCTCTCCCTGCCATCTGTCTTCTCACTGGAAAATTCATCGTTC 760

Y       2661 CCGAGATTAGCAACTATGCGAGTATCCTCTTCATGGCGCTCTTCTCGTCGATTGCAATAA 2720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       761 CCGAGATTAGCAACTATGCGAGTATCCTCTTCATGGCGCTCTTCTCGTCGATTGCAATAA 820

Y       2721 CGGGTATTCTCGAGATGCAATGGGGCAAAGTTGGGATCGATGATTGGTGGAGAAACGAAC 2780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       821 CGGGTATTCTCGAGATGCAATGGGGCAAAGTTGGGATCGATGATTGGTGGAGAAACGAAC 880

Y       2781 AGTTTTGGGTCATTGGAGGTGTTTCTGCGCATCTGTTTGCTCTCTTCCAAGGTCTCCTCA 2840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       881 AGTTTTGGGTCATTGGAGGTGTTTCTGCGCATCTGTTTGCTCTCTTCCAAGGTCTCCTCA 940

Y       2841 AGGTTCTTGCTGGTGTCGACACTAACTTCACAGTCACATCAAAAGCAGCTGATGATGGAG 2900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       941 AGGTTCTTGCTGGTGTCGACACTAACTTCACAGTCACATCAAAAGCAGCTGATGATGGAG 1000

Y       2901 AGTTCTCTGACCTTTACCTCTTCAAATGGACTTCACTTCTCATCCCTCCAATGACTCTAC 2960
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1001 AGTTCTCTGACCTTTACCTCTTCAAATGGACTTCACTTCTCATCCCTCCAATGACTCTAC 1060

Y       2961 TCATCATAAACGTCATTGGAGTCATAGTCGGAGTCTCTGATGCCATCAGCAATGGATACG 3020
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1061 TCATCATAAACGTCATTGGAGTCATAGTCGGAGTCTCTGATGCCATCAGCAATGGATACG 1120

Y       3021 ACTCGTGGGGACCGCTTTTCGGAAGACTGTTCTTTGCACTTTGGGTCATCATTCATCTTT 3080
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1121 ACTCGTGGGGACCGCTTTTCGGAAGACTGTTCTTTGCACTTTGGGTCATCATTCATCTTT 1180

Y       3081 ACCCGTTCCTTAAAGGTTTGCTTGGGAAACAAGATAGAATGCCAACCATTATTGTCGTCT 3140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1181 ACCCGTTCCTTAAAGGTTTGCTTGGGAAACAAGATAGAATGCCAACCATTATTGTCGTCT 1240

Qy      3141 GGTCCATCCTCCTGGCCTCGATTCTTACACTTCTTTGGGTCCGGGTTAATCCGTTTGTGG 3200
             ||||||||||| || ||||||||| ||||||||||||||||||||||||||||||||||
Db      1241 GGTCCATCCTCCTGGCCTCGATTCTTACACTTCTTTGGGTCCGGGTTAATCCGTTTGTGG 1300

Qy      3201 CGAAAGGCGGTCCTATTCTCGAGATCTGTGGTTTAGACTGCTTGTGATTCGATTGACCGG 3260
             |||||| |||| |||||||||||||||| |||||||||||||||||||||||||||||||
Db      1301 CGAAAGGCGGTCCTATTCTCGAGATCTGTGGTTTAGACTGCTTGTGATTCGATTGACCGG 1360

Qy      3261 TGGATGGGTTGGTGAAAAAGGTTTAATTCCCACGGATCAAAGAGAGGTAAGAGAGATATT 3320
             |||||| ||||||||||||||||||||||| |||| |||||||||||||||||||||||
Db      1361 TGGATGGGTTGGTGAAAAAGGTTTAATTCCCACGGATCAAAGAGAGGTAAGAGAGATATT 1420

Qy      3321 GTTTTACCTCTAAAAGACTCCTTCATTGTGTTCATTAGATGATGAAAAATGAAAAGAAAA 3380
             ||| |||| ||| |||||||| |||| |||||||||||||||||||||||||||||||||
Db      1421 GTTTTACCTCTAAAAGACTCCTTCATTGTGTTCATTAGATGATGAAAAATGAAAAGAAAA 1480

Qy      3381 AGAAGATTTAA 3391
             ||||||||| |
Db      1481 AGAAGATTTAA 1491

RESULT 4
AY143957
LOCUS       AY143957                1101 bp    mRNA    linear   PLN 05-OCT-2002
DEFINITION  Arabidopsis thaliana At5g64740/MVP7_7 mRNA, complete cds.
ACCESSION   AY143957
VERSION     AY143957.1  GI:23506072
KEYWORDS    FLI_CDNA.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Cheuk,R., Chen,H., Kim,C.J., Shinn,P., Bowser,L., Carninci,P.,
            Chan,M.M., Chang,C.H., Dale,J.M., Deng,J.M., Hayashizaki,Y.,
            Hsuan,V.W., Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G.,
            Kawai,J., Lam,B., Lee,J.M., Lin,J., Miranda,M., Narusaka,M.,
            Nguyen,M., Palm,C.J., Quach,H.L., Sakurai,T., Satou,M., Seki,M.,
            Southwick,A., Tang,C.C., Toriumi,M., Wallender,E.K., Wong,C.,
            Wu,H.C., Yamada,K., Yu,G., Yuan,S., Shinozaki,K., Davis,R.W.,
            Theologis,A. and Ecker,J.R.
  TITLE     Arabidopsis ORF clones
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1101)
  AUTHORS   Cheuk,R., Chen,H., Kim,C.J., Shinn,P., Bowser,L., Carninci,P.,
            Chan,M.M., Chang,C.H., Dale,J.M., Deng,J.M., Hayashizaki,Y.,
            Hsuan,V.W., Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G.,
            Kawai,J., Lam,B., Lee,J.M., Lin,J., Miranda,M., Narusaka,M.,
            Nguyen,M., Palm,C.J., Quach,H.L., Sakurai,T., Satou,M., Seki,M.,
            Southwick,A., Tang,C.C., Toriumi,M., Wallender,E.K., Wong,C.,
            Wu,H.C., Yamada,K., Yu,G., Yuan,S., Shinozaki,K., Davis,R.W.,
            Theologis,A. and Ecker,J.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-AUG-2002) Salk Institute Genomic Analysis Laboratory
            (SIGnAL), Plant Biology Laboratory, The Salk Institute for
            Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037,
            USA
COMMENT     RIKEN Genomic Sciences Center (GSC) members carried out the
            collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN
            Arabidopsis Full-Length cDNA') : Seki,M., Narusaka,M., Ishida,J.,
            Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J.,
            Hayashizaki,Y. and Shinozaki,K.

            The Salk, Stanford, PGEC (SSP) Consortium members constructed and
            sequenced the pUNI (ORF) clones using the RAFL cDNAs: Cheuk,R.,
            Chen,H., Kim,C.J., Shinn,P., Bahk,J., Bowser,L., Chan,M.M.,
            Chang,E., Dale,J.M., Deng,J.M., Goldsmith,A.D., Jones,T.,
            Karlin-Neumann,G., Lam,B., Lee,J.M., Lin,J., Miranda,M., Nguyen,M.,
            Onodera,C.S., Palm,C.J., Quach,H.L., Southwick,A., Tang,C.C.,
            Toriumi,M., Wu,H.C., Yamada,K., Yamamura, Y., Yu,G., Davis,R.W.,
            Theologis,A., and Ecker,J.R.
```

```
          Cheuk,R. (SSP/Salk) and Seki,M. (RIKEN GSC) contributed equally to
          this work.  Shinozaki,K. (RIKEN GSC) and Ecker,J.R. (SSP/Salk)
          contributed equally to this work as PIs.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
                     /chromosome="5"
                     /clone="U12458"
                     /note="This clone is in pUNI 51
                     ecotype: Columbia"
     CDS             1. .1101
                     /note="cellulose synthase catalytic subunit"
                     /codon_start=1
                     /product="At5g64740/MVP7_7"
                     /protein_id="AAN28896.1"
                     /db_xref="GI:23506073"
                     /translation="MQMKLEKKFGQSPVFVASARMENGGMARNASPACLLKEAIQVIS
                     CGYEDKTEWGKEIGWIYGSVTEDILTGFKMHSHGWRSVYCTPKLAAFEGSAPINLSDR
                     LHQVLRWALGSVEIFLSRHCPIWYGYGGGLKWLERLSYINSVVYPWTSLPLIVYCSLP
                     AICLLTGKFIVPEISNYASILFMALFSSIAITGILEMQWGKVGIDDWWRNEQFWVIGG
                     VSAHLFALFQGLLKVLAGVDTNFTVTSKAADDGEFSDLYLFKWTSLLIPPMTLLIINV
                     IGVIVGVSDAISNGYDSWGPLFGRLFFALWVIIHLYPFLKGLLGKQDRMPTIIVVWSI
                     LLASILTLLWVRVNPFVAKGGPILEICGLDCL"
BASE COUNT      243 a    235 c    272 g    351 t
ORIGIN

  Query Match             24.0%;  Score 827;  DB 8;  Length 1101;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  827;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2421 AAGGATCAGCTCCAATCAATCTTTCGGATCGTCTCCATCAAGTTCTTCGATGGGCGCTTG 2480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        275 AAGGATCAGCTCCAATCAATCTTTCGGATCGTCTCCATCAAGTTCTTCGATGGGCGCTTG 334

Qy       2481 GGTCGGTTGAGATTTTCTTGAGTAGGCATTGTCCTATTTGGTATGGTTATGGAGGTGGGT 2540
             ||||||||||||||||||||||| |||| |||||||||||| |||| ||||||||||||
Db        335 GGTCGGTTGAGATTTTCTTGAGTAGGCATTGTCCTATTTGGTATGGTTATGGAGGTGGGT 394

Qy       2541 TGAAATGGCTTGAGCGGTTGTCCTACATTAACTCTGTGGTTTACCCGTGGACCTCTCTAC 2600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        395 TGAAATGGCTTGAGCGGTTGTCCTACATTAACTCTGTGGTTTACCCGTGGACCTCTCTAC 454

Qy       2601 CGCTCATCGTTTACTGTTCTCTCCCTGCCATCTGTCTTCTCACTGGAAAATTCATCGTTC 2660
             |||| |||||||||||| |||| |||  |||| ||| |||| ||| |||||||||||||
Db        455 CGCTCATCGTTTACTGTTCTCTCCCTGCCATCTGTCTTCTCACTGGAAAATTCATCGTTC 514

Qy       2661 CCGAGATTAGCAACTATGCGA[illegible]TTGCAATAA 2720
             |||||||||||||||||||| [illegible] ||||||||
Db        515 CCGAGATTAGCAACTATGCGA[illegible]TTGCAATAA 574

Qy       2721 CGGGTATTCTCGAGATGCAATGGGGCAAAGTTGGGATCGAT[illegible]GGAGAAACGAAC 2780
             |||||||||||||||||||| |||| |||||||||||||||| [illegible] ||||||||||
Db        575 CGGGTATTCTCGAGATGCAATGGGGCAAAGTTGGGATCGATGATTGGTGGAGAAACGAAC 634

Qy       2781 AGTTTTGGGTCATTGGAGGTGTTTCTGCGCATCTGTTTGCTCTTCTTCCAAGGTCTCCTCA 2840
             |||||||||||||||||| |||||| |||| ||| |||||||||||||||||||||||||||
Db        635 AGTTTTGGGTCATTGGAGGTGTTTCTGCGCATCTGTTTGCTCTTCTTCCAAGGTCTCCTCA 694

Qy       2841 AGGTTCTTGCTGGTGTCGACACTAACTTCACAGTCACATCAAAAGCAGCTGATGATGGAG 2900
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        695 AGGTTCTTGCTGGTGTCGACACTAACTTCACAGTCACATCAAAAGCAGCTGATGATGGAG 754

Qy       2901 AGTTCTCTGACCTTTACCTCTTCAAATGGACTTCACTTCTCATCCCTCCAATGACTCTAC 2960
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        755 AGTTCTCTGACCTTTACCTCTTCAAATGGACTTCACTTCTCATCCCTCCAATGACTCTAC 814

Qy       2961 TCATCATAAACGTCATTGGAGTCATAGTCGGAGTCTCTGATGTCATCAGCAATGGATACG 3020
             |||||||||||||||||||| |||||||||||||||||||||| ||| |||||||||| |
Db        815 TCATCATAAACGTCATTGGAGTCATAGTCGGAGTCTCTGATGTCATCAGCAATGGATACG 874

Qy       3021 ACTCGTGGGGACCGCTTTTCGGAAGACTGTTCTTTGCACTTTGGGTCATCATTCATCTTT 3080
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        875 ACTCGTGGGGACCGCTTTTCGGAAGACTGTTCTTTGCACTTTGGGTCATCATTCATCTTT 934

Qy       3081 ACCCGTTCCTTAAAGGTTTGCTTGGGAAACAAGATAGAATGCCAACCATTATTGTCGTCT 3140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        935 ACCCGTTCCTTAAAGGTTTGCTTGGGAAACAAGATAGAATGCCAACCATTATTGTCGTCT 994

Qy       3141 GGTCCATCCTCCTGGCCTCGATTCTTACACTTCTTTGGGTCCGGGTTAATCCGTTTGTGG 3200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        995 GGTCCATCCTCCTGGCCTCGATTCTTACACTTCTTTGGGTCCGGGTTAATCCGTTTGTGG 1054

Qy       3201 CGAAAGGCGGTCCTATTCTCGAGATCTGTGGTTTAGACTGCTTGTGA 3247
             ||||||||||||||||||||||||||||||||||||||||||||||||
Db       1055 CGAAAGGCGGTCCTATTCTCGAGATCTGTGGTTTAGACTGCTTGTGA 1101


RESULT 5
AB025637
LOCUS       AB025637               39645 bp    DNA     linear   PLN 27-DEC-2000
DEFINITION  Arabidopsis thaliana genomic DNA, chromosome 5, P1 clone:MVP7.
ACCESSION   AB025637 BA000015
VERSION     AB025637.1  GI:4589443
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (sites)
  AUTHORS   Kaneko,T., Katoh,T., Asamizu,E., Sato,S., Nakamura,Y., Kotani,H.
            and Tabata,S.
  TITLE     Structural analysis of Arabidopsis thaliana chromosome 5. XI
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 39645)
  AUTHORS   Nakamura,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-APR-1999) Yasukazu Nakamura, Kazusa DNA Research
            Institute, Department of Plant Gene Research; 1532-3, Yana,
            Kisarazu, Chiba 292-0812, Japan (E-mail:ynakamu@kazusa.or.jp,
            Tel:81-438-52-3935, Fax:81-438-52-3934)
COMMENT     Address for correspondence: kaos@kazusa.or.jp
            For the latest information on annotation of this clone, please see
            http://www.kazusa.or.jp/kaos/cgi-bin/agd_graph.cgi?c=MVP7
            Genes with similarity to proteins in the databases are described in
            'product' or 'note' qualifiers.  Genes that have no significant
            protein similarity are described as 'unknown protein'.
            The software programs used to predict genes include: Grail
            (Informatics Group, Oak Ridge National Laboratory,
            http://compbio.ornl.gov/Grail-1.3/),
            GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),
            NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of
            Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and
            SplicePredictor (Volker Brendel, Stanford University,
            http://gremlin1.zool.iastate.edu/cgi-bin/sp.cgi).
            Genes encoding tRNAs are predicted by tRNAscan-SE
            (Sean Eddy, Washington University School of Medicine, St. Louis,
            http://genome.wustl.edu/eddy/tRNAscan-SE/).
            This sequence may not be the entire insert of this clone. It may be
            shorter because we remove overlaps between neighboring submissions.
            The 5' clone is MUB3 and the 3' clone is MXK3.
FEATURES             Location/Qualifiers
     source          1. .39645
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="5"
                     /clone="MVP7"
                     /clone_lib="Mitsui P1"
     CDS             complement(join(2170. .2317,2399. .2550,2646. .2804,
                     2919. .3180,3283. .3399,3505. .3567,3647. .3825))
                     /note="gene_id:MVP7.2"
```

/codon_start=1
/evidence=not_experimental
/product="nodulin-like protein"
/protein_id="BAB10303.1"
/db_xref="GI:10177201"
/translation="MDMESKKPYLMVTIIQVIYTIMFLISKAVFNGGMNTFVFVFYRQ
AFATIFLAPLAFFFERKSAPPLSFVTFIKIFMLSLFGVTLSLDLNGIALSYTSATLAA
ATTASLPAITFFLALLFGMERLKVKSIQGTAKLVGITVCMGGVIILAIYKGPLLKLPL
CPHFYHGQEHPHRNNPGHVSGGSTSWLKGCVLMITSNILWGLWLVLQGRVLKVYPSKL
YFTTLHCLLSSIQSFVIAIALERDISAWKLGWNLRLVAVIYCGFIVTGVAYYLQSWVI
EKRGPVFLSMFTPLSLLFTLLSSAILLCEIISLGSIVGGLLLIIGLYCVLWGKSREEK
NSGDDKIDLQKENDVVCNEVKVVIS"
CDS join(6348. .6738,6847. .8146,8238. .9072)
/note="emb|CAB71880.1
gene_id:MVP7.3
similar to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB10304.1"
/db_xref="GI:10177202"
/translation="MQFGSSKTLFFRVAGFSSTSVFKPRRAISVPVGNFSSSSSISAH
YTSLKRSQYEEDSRSVRVPVWWDFENCHLPSGANVFKLAQTITSAVRICGIKGPITIT
AYGDMIQLSRTNQEALFATGINLTHVPQGGKNSTDRSLITEIMCWVSQNPPPAHLFLI
SSDSDFANVLHRLRMRNYNILLACYEETTLGVLCSAASIMWDWDALVRGQNPTAKHFN
QPPDGPYHSWYGHYTTPLLDPFATSTNNKQISSTSVKTVELLELGSSKSRRPIPNKVV
KQIGLILRWYPKGAAITELREQLRKRKVLLDRDFYGYKSFSRFLLSMRNILQVVPLGD
GMFSIHAVTQGMDNKALLPKVSCENHAVVSVEKMCQNMKQNDKDVKEESHQLQESSQE
FVQVMKLMDVKAKEEPVKANQLAITAVDDVSSFEEKDGFLKKLNRLLFGSPEMELEHL
QERKHISGNGVLGEGKVVDKDLESQIASSTSSESAEEVKVDNAVGNGKSKSPGLTCRL
LKRFKFSWGRYTELSNAAATGPQVDDVFVKDSFWNDVESFINSPRGFVAVSHSRSRET
MAKNLKEEGPSSLKPLDVPKMLDLISMLISEKKWIQENPSDALPFRVTRFTEESSCRS
NPRTTDGLRAIFVNMSESLCDGANGDKNSTNVGMSQKPKERSRSKVIADCHKLIKKIT
EENAGGYSITKFKKAFLEKFGYRLEYRKFGFSKLQSLIEMMPEARIESGHIVTSSTPV
PCESDSSFEDLGPVSKKIHENESSVSEGEDYDSEMEEKASSKQSGGERKKKEDETESD
LLQILGSWDTDKKPAKTFGEDKLVEGILMSLRKKPSGDSRIQD"
CDS complement(9326. .9793)
/note="gb|AAD20394.1
gene_id:MVP7.4
similar to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB10305.1"
/db_xref="GI:10177203"
/translation="MATKSTSKPLLLSFLMMSYLISTFHVITVAEGRTLQFTKMATDH
SGAGNLMDCWNAGLELKSCTDEIVKFFLSQTGTSEPPVKGGIDKDCCGAIGLVVKDCW
SVMFTSLGLTTMEGNNLREYCEFQAEKSELSPSPAPETLALSPVEITYPGLDY"
CDS join(10126. .10334,10710. .10812,10911. .11018,11125. .11234,
11323. .11430,11525. .11570,11688. .11819,11866. .12081)
/note="contains similarity to unknown protein
gb|AAF44992.1
gene_id:MVP7.5"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB10306.1"
/db_xref="GI:10177204"
/translation="MSATELPTKEAHILKGHEGAVLAARFNGDGNYALTCGKDRTIRL
WNPHRGILIKTYKSHGREVRDVHVTSDNAKFCSCGGDRQVYYWDVSTGRVIRKFRGHD
GEVNAVKFNDSSSVVVSAGFDRSLRVWDCRSHSVEPVQIIDTFLDTVMSVVLTKTEII
GGSVDGTVRTFDMRIGREMSDNLGQPVNCISISNDGNCVLAGCLDSTLRLLDRTTGEL
LQVYKGHISKSFKTDCCLTNSDAHVIGGSEDGLVFFWDLVDAKVLSKFRAHDLVNSID
VDFYLRLLKFFVAGDKRELPPEGRLYVDFFGRWYNSCLEKMKKQQILTSIDFVIYVSN
SPFFFYYRL"
tRNA complement(12371. .12445)
/product="tRNA-His"
/note="codon recognized: CAC; gene_id:MVP7.6"
/evidence=not_experimental
CDS join(18535. .18615,19010. .19205,19294. .19496,19580. .19857,
19940. .20206,20290. .20635,20823. .20960,21047. .21172,
21284. .21496,21580. .21856,21942. .22138,22221. .22571,
22732. .23313)
/note="gene_id:MVP7.7"
/codon_start=1
/evidence=not_experimental
/product="cellulose synthase catalytic subunit"
/protein_id="BAB10307.1"
/db_xref="GI:10177205"
/translation="MNTGGPLIAGSHNRNEFVLINADENARIRSVQELSGQTCQICRD
EIELTVEGEPFVACNECAFPVCRPCYEYERREGNQACPQCKTRFKRLKGSPRVEGDEE
EDDIDDVENEFEYGNKGIGFDQVSEGMSISRRNSGFPQSDLDSAPPGSQIPLLTYGDE
DVEISSERHALIVPPSLGGHGNRVHPVSLSDPTVAAHPRPMVPQKDLAVYGYGSVAWK
DRMEEWKPKQNEKLQVVRHEGDPDPEDGDDADFPMMDEGRQPLSRKIPIKSSKINPYR
MLIVLRLVILGLFFHYRILHPVKDAYALWLISVICEIWFAVSWVLDQFPKWYPIERET
YLDRLSLRYEKEGKPSGLSPVDVFVSTVDPLKEPPLITANTVLSILAVDYPVDKVACY
VSDDGAAMLTFEALSETAEFARKWVPFCKKYCIEPRAPEWYFCHKMDYLKNKVHPAFV
RERRAMKRDYEEFKVRINALVATAQKVPEDGWTMQDGTPWPGNSVRDHPGMIQVFLGS
DGVRDVENNELPRLVYVSREKRPGFDHHKKAGAMNSLIRVSGVLSNAPYLLNVDCDHY
INNSKALREAMCFMMDPQSGKKICYVQFPQRFDGIDRHDRYSNRNVVFFDINMKGLDG
LQGPIYVGTGCVFRRQALYGFDAPKKKKGPRKTCNCWPKWCLLCFGSRKNRKAKTVAA
DKKKKNFEASKQTHALENIEEGRVTKQSNVEQSTEAMQMKLEKKFGQSPVFVASARME
NGGMAPNASPACLLKEAIQVISCGYEDKTEWGKEIGWIYGSVTEDILTGFKMHSHGWR
SVYCTPKEAAFKGSAPINLTDRLHQVLRWALGSVEIFLSRHCPIWYGYGGGLKWLERL
AYINLSVYPWTSLPLIVYCTLPAICLLTGKFIVPEISNYASILFMALFSSIAITGILE
MRWSGVSIEDWWRNEQFWVIGGVSAHLFAVFQGLLKVLAGVDTNFTVTSKAADDGEFS
ELYLFKWTSLLIPPMTLLIINVIGVVVGVSDAISNGYDSWGPLFGRLFFALWVIIHLY
PFLKGLLGRQDRMPTIVVVWSILLASIFSLLWVRVNPFVAKGGPILEICGLDCL"
CDS join(28659. .29097,29900. .30636)
/note="contains similarity to transcription factor
gene_id:MVP7.8"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB10308.1"
/db_xref="GI:10177206"
/translation="MCVLKVANQEDNVGKKAESIRDDDHRTLSEIDQWLYLFAAEDDH
HRHSFPTQQPPPSSSSSSLISGFSREMEMSAIVSALTHVVAGNVPQHQQGGGEGSGEG
TSNSSSNSGQKRRREVEEGGAKAVKAANTLTVDQYFSGGSSTSKVREASSNMSGPGPT
YEYTTTATASSETSSFSGDQPRRRYKGVRQRPWGKWAAEIRDPFKAARVWLGTFDNAE
SAARAYDEAALRFRGNKAKLNFPENVKLVRPASTEAQPVHQTAAQRPTQSRNSGSTTT
LLPIRFASNQSVHSQFLMQSYNLSYSEMARQQQQFQQHHQQSLDLYDQMSFPLRFGHT
GGSMMQSTSSSSSHSRPLFSPAAVQPPPESASETGYLQDIQWPSDKTSNNYNNSPSS"
CDS complement(join(31307. .31225,31303. .31380,31479. .31632,
31734. .31846,31941. .32040,32184. .32380,32485. .32541,
32640. .32744,32857. .32964,33125. .33319,33423. .33552,
33700. .33833))
/note="gene_id:MVP7.9"
/codon_start=1
/evidence=not_experimental
/product="proteasome regulatory subunit-like"
/protein_id="BAB10309.1"
/db_xref="GI:10177207"
/translation="MGLKKLTQMGHMDLRFSFITNGSYSFVHNILPNSYLGMTLQATG
RGWGFDANILSLSEPLEVKSVSSPESFTNSSSLSRSIDSRRLAMEESRQLESSIDRLL
NEEKQMRLAENVAGTRKAATEILKLCFEAKDWKLLNEQILNLSKKRGQLKQAVQSMVQ
QAMQYIDQTLDIETRIELIKTLNNVAAGKIYVEIERARLTKMLAKIKEEQGLIAEAAD
LMQEVAVETFGAMAKTEKIAFILEQVRLCLDQKDFVRAQILSRKINPRVFDADTTKEK
KKPKEGENMVEEAPADIPTLLVLKRIYYELMIRYYSHNNEYIEICRSYKAIYDIPSVK
ENPEQWIPVLRKICWFLALAPHDPMQSSLLNATLEDKKLSEIPDFKMLLKQIVTMEVI
QWTSLWNKYKDEFENEKNMIGGSLGDKAGEDLKLRIIEHNILVVSKYYSRITFKRLAE
LLCLTTEEAEKHLSEMVVSKALIAKIDRFSGIICFQIVKDSNEILNSWATNLEKLLDL
VEKSCHQTHKETMVHKVALRA"
CDS complement(join(34661. .34824,34990. .35065))
/note="gene_id:MVP7.10
unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB10310.1"
/db_xref="GI:10177208"
/translation="MAIRVSHKSFLVALLLILFISSPTQARSLREVVRNRTLLVVEKS
QESRKIRHEGGGSDVDGLMMDYNSANKKRPIHNR"
CDS complement(join(37680. .37967,38133. .38213,38777. .38935))
/note="gene_id:MVP7.11
pir||T04031
similar to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB10311.1"
/db_xref="GI:10177209"

```
  Query Match             22.4%;  Score 773;  DB 8;  Length 39645;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  773;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2664 AGATTAGCAACTATGCGAGTATCCTCTTCATGGCGCTCTTCTCGTCGATTGCAATAACGG 2723
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      22730 AGATTAGCAACTATGCGAGTATCCTCTTCATGGCGCTCTTCTCGTCGATTGCAATAACGG 22789

Qy       2724 GTATTCTCGAGATGCAATGGGGCAAAGTTGGGATCGATGATTGGTGGAGAAACGAACAGT 2783
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      22790 GTATTCTCGAGATGCAATGGGGCAAAGTTGGGATCGATGATTGGTGGAGAAACGAACAGT 22849

Qy       2784 TTTGGGTCATTGGAGGTGTTTCTGCGCATCTGTTTGCTCTCTTCCAAGGTCTCCTCAAGG 2843
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      22850 TTTGGGTCATTGGAGGTGTTTCTGCGCATCTGTTTGCTCTCTTCCAAGGTCTCCTCAAGG 22909

Qy       2844 TTCTTGCTGGTGTCGACACTAACTTCACAGTCACATCAAAAGCAGCTGATGATGGAGAGT 2903
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      22910 TTCTTGCTGGTGTCGACACTAACTTCACAGTCACATCAAAAGCAGCTGATGATGGAGAGT 22969

Qy       2904 TCTCTGACCTTTACCTCTTCAAATGGACTTCACTTCTCATCCCTCCAATGACTCTACTCA 2963
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      22970 TCTCTGACCTTTACCTCTTCAAATGGACTTCACTTCTCATCCCTCCAATGACTCTACTCA 23029

Qy       2964 TCATAAACGTCATTGGAGTCATAGTCGGAGTCTCTGATGCCATCAGCAATGGATACGACT 3023
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      23030 TCATAAACGTCATTGGAGTCATAGTCGGAGTCTCTGATGCCATCAGCAATGGATACGACT 23089

Qy       3024 CGTGGGGACCGCTTTTCGGAAGACTGTTCTTTGCACTTTGGGTCATCATTCATCTTTACC 3083
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      23090 CGTGGGGACCGCTTTTCGGAAGACTGTTCTTTGCACTTTGGGTCATCATTCATCTTTACC 23149

Qy       3084 CGTTCCTTAAAGGTTTGCTTGGGAAACAAGATAGAATGCCAACCATTATTGTCGTCTGGT 3143
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      23150 CGTTCCTTAAAGGTTTGCTTGGGAAACAAGATAGAATGCCAACCATTATTGTCGTCTGGT 23209

Qy       3144 CCATCCTCCTGGCCTCGATTCTTACACTTCTTTGGGTCCGGTTAATCCGTTTGTGGCGA 3203
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      23210 CCATCCTCCTGGCCTCGATTCTTACACTTCTTTGGGTCCGGTTAATCCGTTTGTGGCGA 23269

Qy       3204 AAGGCGGTCCTATTCTCGAGATCTGTGGTTTAGACTGCTTGTGATTCGATTGACCGGTGG 3263
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      23270 AAGGCGGTCCTATTCTCGAGATCTGTGGTTTAGACTGCTTGTGATTCGATTGACCGGTGG 23329

Qy       3264 ATGGGTTGGTGAAAAAGGTTTAATTCCCACGGATCAAAGAGAGGTAAGAGAGATATTGTT 3323
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      23330 ATGGGTTGGTGAAAAAGGTTTAATTCCCACGGATCAAAGAGAGGTAAGAGAGATATTGTT 23389

Qy       3324 TTACCTCTAAAAGACTCCTTCATTGTGTTCATTAGATGATCAAAATGAAAAGAAAAAGA 3383
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      23390 TTACCTCTAAAAGACTCCTTCATTGTGTTCATTAGATGATCAAAATGAAAAGAAAAAGA 23449

Qy       3384 AGATTTAATTTTGTTACGAGAATTGTTATTTTTGCAAGAATTTGTTGTAGATA 3436
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      23450 AGATTTAATTTTGTTACGAGAATTGTTATTTTTGCAAGAATTTGTTGTAGATA 23502


RESULT 6
AB016893
LOCUS       AB016893               82390 bp    DNA     linear   PLN 27-DEC-2000
DEFINITION  Arabidopsis thaliana genomic DNA, chromosome 5, P1 clone:MYH9.
ACCESSION   AB016893 BA000015
VERSION     AB016893.1  GI:3449334
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (sites)
  AUTHORS   Asamizu,E., Sato,S., Kaneko,T., Nakamura,Y., Kotani,H., Miyajima,N.
            and Tabata,S.
  TITLE     Structural analysis of Arabidopsis thaliana chromosome 5. VIII.
            Sequence features of the regions of 1,081,958 bp covered by
            seventeen physically assigned P1 and TAC clones
  JOURNAL   DNA Res. 5 (6), 379-391 (1998)
  MEDLINE   99156233
   PUBMED   10048488
REFERENCE   2  (bases 1 to 82390)
  AUTHORS   Nakamura,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-AUG-1998) Yasukazu Nakamura, Kazusa DNA Research
            Institute, Department of Plant Gene Research; 1532-3, Yana,
            Kisarazu, Chiba 292-0812, Japan (E-mail:ynakamu@kazusa.or.jp,
            Tel:81-438-52-3935, Fax:81-438-52-3934)
COMMENT     Address for correspondence: kaos@kazusa.or.jp
            For the latest information on annotation of this clone, please see
            http://www.kazusa.or.jp/kaos/cgi-bin/agd_graph.cgi?c=MYH9
            Genes with similarity to proteins in the databases are described in
            'product' or 'note' qualifiers.  Genes that have no significant
            protein similarity are described as 'unknown protein'.
            The software programs used to predict genes include: Grail
            (Informatics Group, Oak Ridge National Laboratory,
            http://compbio.ornl.gov/Grail-1.3/),
            GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),
            NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of
            Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and
            SplicePredictor (Volker Brendel, Stanford University,
            http://gremlin1.zool.iastate.edu/cgi-bin/sp.cgi).
            Genes encoding tRNAs are predicted by tRNAscan-SE
            (Sean Eddy, Washington University School of Medicine, St. Louis,
            http://genome.wustl.edu/eddy/tRNAscan-SE/).
            This sequence may not be the entire insert of this clone. It may be
            shorter because we remove overlaps between neighboring submissions.
            The 5' clone is MTH16.
FEATURES             Location/Qualifiers
     source          1. .82390
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="5"
                     /clone="MYH9"
                     /clone_lib="Mitsui P1"
     CDS             join(5573. .5632,5727. .6120,6207. .6820,6919. .6984)
                     /note="gene_id:MYH9.2
                     sp|P53492"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="ACTIN 2/7"
                     /protein_id="BAB09402.1"
                     /db_xref="GI:9758959"
                     /translation="MADGEDIQPLVCDNGTGMVKAGFAGDDAPRAVFPSIVGRPRHTG
                     VMVGMGQKDAYVGDEAQSKRGILTLKYPIEHGIVSNWDDMEKIWHHTFYNELRVAPEE
                     HPVLLTEAPLNPKANREKMTQIMFETFNVPAMYVAIQAVLSLYASGRTTGIVLDSGDG
                     VSHTVPIYEGYALPHAILRLDLAGRDLTDSLMKILTERGYMFTTTAEREIVRDIKEKL
                     AYVALDYEQELETAKSSSSVEKNYELPDGQVITIGAERFRCPEVLFQPSLIGMEAPGI
                     HETTYNSIMKCDVDIRKDLYGNIVLSGGSTMFPGIADRMSKEITALAPSSMKIKVVAP
                     PERKYSVWIGGSILASLSTFQQMWISKSEYDESGPSIVHRKCF"
     CDS             complement(join(9228. .9413,9491. .9729,9859. .10144))
                     /note="gene_id:MYH9.3
                     pir||S75332
                     similar to unknown protein"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="BAB09403.1"
                     /db_xref="GI:9758960"
                     /translation="MTSNLFQPPSMAASRGAISRRTGNVKVLVSFTSSNGKTLSFSDN
                     SFRLRPMFIGKVTEQSSCSSPNEQQQDEEQEQEQEEITVSHIKEELYEALKGINRGIF
                     GVKSDKKTEIEGLVKLLECRNPTPEPTGELDKIGGCWKLIYSTITVLGSKRTKLGLRD
                     FVSLGDLLQQIDIAQGKTVHVLKFDVRGLNLLDGEFRIVASFKISSKSVHYRFLDVFR
                     FFRFLEKSRFDGIAFVCL"
     CDS             complement(join(10580. .10778,10867. .10889,11474. .11533))
                     /note="gene_id:MYH9.4
                     unknown protein"
```

/codon_start=1
/evidence=not_experimental
/protein_id="BAB09404.1"
/db_xref="GI:9758961"
/translation="MVTKEQVEASLTSKLKPIHLEVIDISGGCGSSFEVEVVSEQFEG
KRLLERHRMVNAALEEEMKEIHALSIKKAQTPQQWKPPSQDSATLTKDA"
CDS join(11791. .13676,13846. .14734)
/note="emb|CAB71880.1
gene_id:MYH9.5
similar to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB09405.1"
/db_xref="GI:9758962"
/translation="MKFGSLPSKSLLFRLSSSSTTQSHRRTITLFITNIHSSPFSTTT
TSGSNFVSGSSHSPSRRPQQDEESRSVRVSVWWDFLSCNLPVDTNVYKVAQSITAAIR
NSGIKGPITITAFGDVLQLPRSNQDALSATGISLTHVPNGGKNSADRSLITDLMCWVS
QNPPPAHLLLISSDKEFASVLHRLRMNNYNILLASKSSAPGVLCSAASIMWDWDALIK
GECVTGKHFNQPPDGPYNSWYGHYRIPLLDPFAIATTTEQSSSSVKIEELSESVNSNA
VVNLRPIPKEVVDKIRSIVSLYPKGAAITELRAELSKSNLAIDKDFYGHKKFSKFLLS
MPDILQVTTVSEGLFMIRAVTEKKPPMRLDSSPRLSTAVDQKIKDKETVNAPSPKLIS
DVELAAERRRDGLLGKKQEKVLESDKIVKEESSESSQDPILVGQKDVKANVKPVETNQ
VALVAWSDSSSMEDGFFQKLKRLWYGSPEMELEHLPEMKSVSGSGDTHREDNDLKSSSQ
GSGNKHKEDKDLKSSSQGTVPMSQISPSFVAESVDEVKVGADDVDSKDKDASPGFLGR
LLKSFKFWGKNTKSSKDSSGNQELVNVDSQVQDIFAKEYFWSDIESFINSPRGFAIVS
HSRTREVMAKNVQEEGPSCLRLLDESSMLHLVTLLISEKKWIEETPSSSLPFRIIKGS
SPGHRHASNGLSSIFSDSSKSQWQKQNGEKSGKNVAHAGVSVGSVDRKELERYKSNAI
ADCQKMIKKITEEHPEGYSLIRFRKDFLEEYGYHLAVDKLGYDNLQSLIRVMHGVRIA
SGYIFPSTPSPNAKSKEDDSDLAFAELGPVSDTTTTHPTTKKLPVYEPSLSEDEEDSG
SERDNPEKKKQQMMSKEGKESSLLQILDSYYTNKDGELKENPVERKLVSNGRKQKPTK
TYSFVKDSEV"
CDS complement(join(16252. .16364,16457. .16535,16624. .16763,
16849. .16882,16958. .17034,17092. .17165,17245. .17327,
17416. .17484,17569. .17985))
/note="gb|AAF19567.1
gene_id:MYH9.6
similar to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB09406.1"
/db_xref="GI:9758963"
/translation="MDLDDPRSVMDNAGVDVWTFIDTAILVASLDYGQELKRRRDNIV
ERLYATSMANKCRNCDFGGGGNVTEAAIGRVNNGRVHEETEEEDEEGVTAAAEEEVRE
KSVNVEDDDDFDPFAGLFDDEQKSIVEIKEKLEDPDLSEESLVELLQNLEDMDITFQA
LQETDIGRHVNRVRKHPSNNVRRLAKQLVKKWKETVDEWVKFNQPGDLEPPSLIEQSI
SSNVADEDSPVQKALHNGSRQQVPDFGYSPVPQNGYSSSSKNSNITEPERKPRPVAPQ
PRRESPSPAKPSRPSPSQQTIPRDKEHKEVDFDTARKRLQQNYRQAENAKKQRTIQVM
DIHDIPKPKKGGFFPRKGGSSQGGRHW"
CDS join(19579. .19581,19681. .19764,19841. .19915,20012. .20114,
20275. .20420,20500. .20571,20698. .20770,21023. .21069,
21259. .21330,21458. .21583,21628. .21726,21806. .21847,
21980. .22096,22189. .22271,22481. .22547,22675. .22746,
22846. .22896,22986. .23045,23123. .23497)
/note="gene_id:MYH9.7
unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB09407.1"
/db_xref="GI:9758964"
/translation="MDAFRDAILQRAPIETFALKTVQHFIQPQKQTKLAQDENQMLEN
MLRTLLQELVAAAAQSGEQIMQYGQLIDDDDDDDDIHGQIPHLLDVVLYLCEKEHVEG
GMIFQLLEDLTEMSTMKNCKDVFGYIESKQDILGKQELFARGKLVMLRTCNQLLRRLS
KANDVVFCGRILMFLAHFFPLSERSGISVDFNFYKTFWSLQEYFCNPASLTSASTKWQ
KFSSSLAVVLNTFDAQPLSEEEGEANSLEEEAATFNIKYLTSSKLMGLEFTHFFFYLS
PQLKDSSFRRHILLQCLIMFDYLRAPGKNDKDLPSETMKEELKSCEDRVKKLLEITPP
KGKEFLRAVEHILERENKWVWWKRDGCPPFEKQPIDKKSPNAGQKKRRQRWRLGNKEL
SQLWRWADQNPNALTDSQRVRTPDIADYWKPLAEDMDPSAGIEDEYHHKNNRVYCWKG
LRFTARQDLEGFSRFTEMGIEGVVPVELLPPEVRSKYQAKPNEKAKRAKKEETKGGSH
ETEGNQIGVSNSEAEAEGGRGDAETMESDAIADTPTPEEQQRLGGSDTENGQEAGQIE
DGETEEAGLMDTDLDHPPMPVS"
CDS join(26120. .26200,26584. .26779,26858. .27033,27111. .27207,
27298. .27469,27552. .28164,28253. .28390,28491. .28616,
28723. .28935,29029. .29293,29371. .29570,29669. .30019,
30157. .30738)
/note="gene_id:MYH9.8"
/codon_start=1
/evidence=not_experimental
/product="cellulose synthase catalytic subunit"
/protein_id="BAB09408.1"
/db_xref="GI:9758965"
/translation="MNTGGRLIAGSHNRNEFVLINADESARIRSVEELSGQTCQICGD
EIELSVDGSFVACNECAFPVCRPCYEYERREGNQSCPQCKTRYKRIKGSPRVEGDEE
DDGIDDLEFEFDYSRSGLESETFSRRNSEFDLASAPPGSQIPLLTYGEEDVEISSDSH
ALIVSPSPSGHIHRVHQPHFPDPAAHPRPMVPQKDLAVYGYGSVAWKDRMEEWKRKQNE
KYQVVKHCNGDSSLGDGDDADLPMMDEGRQPLSRKVPIKSSKINPYRMLIVLRLVILGL
FFHYRITHPVNDAYALWLISVICEIWFAVSWVLDQFPKWYPIERETYLDRLSLRYEKE
GKPSELAPVDVFVSTVDPMKEPPLITANTVLSILAVDYPVDRVACYVSDDGAAMLTFE
ALSETAEFARKWVPFCKKYTIEPRAPEWYFCHKMDYLKNKVHPAFVRERRAMKRDYEE
FKVKINALVAKAQKVPEEGWTMQDGTPWPGNNVRDHPGMIQVFLGNNGVRDVENNELP
RLVYVSREKRPGFDHHKKAGAMNSLIRVSGVLSNAPYLLNVDCDHYINNSKALREAMC
FMMDPLLGKKICYVQFPQRFDGIDRSDRYSNRNVVFFDINMKGLDGLQGPIYVGTGCV
FRRQALYGYAPVKEKKRTKRMYSTKVKWPGKTLFCCGLRKNRKSKTTDKKKKNREASKQIH
ALENIEEGVEFMNDAAKMGNMVAAGIEVNSKFGQSPVFVASAGMENGGLARNASPASLM
FAIQVLTGYEDKTEWGKEIGWIYGSVTEDILTGFKMHSHGWRSVYCTPKIPAFKGSA
PINLSDRLHQVLRWALGSVEIFLSRHCPIWYGYGGGLKWLERLSYINSVVYPWTSIPL
LVYCSLPAICLLTGKFIVPEISNYASIWFMALFGSIAVTGILEMQWGKVGIDDWWRNE
QFWVIGGVSAHLFALFQGLLKVLAGVETNFTVTSKAADDGEFSELYIFKWTSLLIPPT
TLLIINVIGVIVGISDAISNGYDSWGPLFGRLFFAFWVILHLYPFLKGLLGKQDRMPT
IILVWSILLASILTLLWVRVNPFVAKGGPILEICGLDCL"
CDS complement(join(34253. .34634,34723. .34811,34914. .35094,
35194. .35258,35344. .35427,35521. .35598,35679. .35765,
35880. .36093,36593. .36748,37145. .37201,37306. .37378,
37495. .37943))
/note="gene_id:MYH9.9"
/codon_start=1
/evidence=not_experimental
/product="splicing factor-like protein"
/protein_id="BAB09409.1"
/db_xref="GI:9758966"
/translation="MEFDEYEYLEKTVEEGAGSNKKKEESGNERSERSYRKREGGERQ
GEEGGEEERVSSRSKKSRGDGERNGGGKRDRERERHRSSRDKDRERDKVREGSRDKES
DRERSSKERDRSDRDKPRDRERREREKRSSSRSRREEKEPEVVERGSRRHRDKKDEPE
ADPERDQRTVFAYQMPLKATERDVYEFFSKAGKVRDVRLIMDRNSRRSKGVGYIEFYD
VMSVPMAIALSGQLFLGQPVMVKPSEAEKNLAQSNSTTVGGTGPADRKLYVGNLHFNM

Query Match 1.5%; Score 50; DB 8; Length 82390;
Best Local Similarity 100.0%; Pred. No. 1.6e 15;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1697 AATAGCAAAGCTCTTAGAGAAGCAATGTGTTTCATGATGGATCCTCAGTC 1746
||||||||||||||||||||||||||||||||||||||||||||||||||
Db 28792 AATAGCAAAGCTCTTAGAGAAGCAATGTGTTTCATGATGGATCCTCAGTC 28841

RESULT 7
AC007019
LOCUS AC007019 107848 bp DNA linear PLN 11-MAR-2002
DEFINITION Arabidopsis thaliana chromosome 2 clone F7D8 map mi238, complete sequence.
ACCESSION AC007019
VERSION AC007019.5 GI:[illegible]198003
KEYWORDS HTG.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 107848)
AUTHORS Lin,X., Kaul,S., Shea,T.P., Fujii,C.Y., Shen,M., VanAken,S.E., Barnstead,M.E., Mason,T.M., Bowman,C.L., Ronning,C.M., Benito,M.-I., Carrera,A.J., Creasy,T.H., Buell,C.R., Town,C.D., Nierman,W.C., Fraser,C.M. and Venter,J.C.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 107848)
AUTHORS Lin,X.
TITLE Direct Submission

```
  JOURNAL   Submitted (09-MAR-2000) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
REFERENCE   3  (bases 1 to 107848)
  AUTHORS   Town,C.D. and Kaul,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-FEB-2002) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
COMMENT     On Apr 18, 2002 this sequence version replaced gi:6598650.
            On Oct 8, 1997 this sequence version replaced gi:2443866.
             We have determined that YAC YUP8H12 is chimaeric, and is comprised
             of two distinct genomic EcoRI fragments from chromosome I. This
             submission contains the sequence from the EcoRI site at position 1

             (right end) to position 101918 of our previous Phase II
             submission YUP8H12 accession number AC000098. This fragment maps
            at
             the bottom of the right arm of the chromosome between the ATHATPAS
             and m532 markers.
FEATURES             Location/Qualifiers
     source          1. .107848
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /cultivar="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="2"
                     /map="mi238"
                     /clone="F7D8"
     misc_feature    complement(1. .6910)
                     /note="overlap with BAC clone F2G1 (AC007119:1. .6910)."
     gene            complement(1037. .1520)
                     /gene="At2g21690"
                     /note="synonym: F7D8.1"
     mRNA            complement(join(<1037. .1279,1410. .>1520))
                     /gene="At2g21690"
     CDS             complement(join(1037. .1279,1410. .1520))
                     /gene="At2g21690"
                     /codon_start=1
                     /product="putative glycine-rich RNA binding protein"
                     /protein_id="AAD20390.1"
                     /db_xref="GI:4417265"
                     /translation="MATNVEYTCFVRGLDQDTDEKDLTDIFSKFGNVIDSKIIYDRDT
                     GKSRRFGFVTFEEEKSMTDAIMIMDVEESRSKCVNVGSITVEVARQRRKNRSAEFALE
                     LVRLINEKQKKMAGL"
     gene            complement(1586. .1657)
                     /gene="At2g21700"
     tRNA            complement(1586. .1657)
                     /gene="At2g21700"
                     /product="tRNA-Gln"
     gene            1809. .4230
                     /gene="At2g21710"
                     /note="synonym: F7D8.2 predicted by genscan"
     mRNA            join(<1809. .2464,2596. .3319,3388. .3492,3622. .3896,
                     3969. .>4230)
                     /gene="At2g21710"
     CDS             join(1809. .2464,2596. .3319,3388. .3492,3622. .3896,
                     3969. .4230)
                     /gene="At2g21710"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAD20391.1"
                     /db_xref="GI:4417266"
                     /translation="MLLHCNVSYYTSTFSFISSSLRRQDNADDSQDTVIRRRHNARSI
                     SLYIRHNRDLKLNKNPNRSQETFVPPPPPPRRDLDGENRSKLLELSLVTRRTPQFPGS
                     IYAQSASDADIASSLPSLRNFLGSDGDDDGESEREMIVKALEIRRKVTKEIIKESLVR
                     KGRFGITYATNVTDRLGDFVDHVMIQAAALKRLPEFSESRFNLRARTVIEDSNFVPLV
                     RWLKHHELSYNRIAKIICMSKGNLDSIRIMIEWLKSIHVKGEFIAVAFLRSGDNILQR
                     NREELNEIVEYLESNGVRRDWMGYVVGRCPELLSFSMEEVKSRVDFFLKMGMNQNDFG
                     TMVYDYPKIIGFFSFQVMEKKVLKALENTPALRLSFKFIIVLLVLNQINYLKEFGLST
                     EEVGRLLAYKPHLMGCSIEERWKPLVKYFYYLGIPKEGMKRILVVKPILYCIDLEKTI
                     APKVVELRYNVRFLQEMGIPNEAIGNMLVKFPSLLTNSLYKKIRPVIFLLTRAGVTQK
                     DIGKVIAMDPALLGCSIGTKLEPNMRYYISLGIRFYQLGEMIADFPMLLRYNVDNLRP
                     KYRYLRRTMIRPLQDLIEPPRFFSYSLERRIIPRHTIMVENRVNPKLRYMLACTDEEF
                     ERRVRDKVERRERFEAGLDSEDSQPSDENISDQETAFSDEAEEEEDLTE"
     gene            4619. .7725
                     /gene="At2g21720"
                     /note="synonym: F7D8.4; predicted by genscan"
     mRNA            join(<4619. .4857,5070. .5304,5383. .5484,5578. .5747,
                     5857. .5990,6067. .6244,6329. .6896,7147. .7413,7507. .>7725)
                     /gene="At2g21720"
     CDS             join(4619. .4857,5070. .5304,5383. .5484,5578. .5747,
                     5857. .5990,6067. .6244,6329. .6896,7147. .7413,7507. .7725)
                     /gene="At2g21720"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAD20392.1"
                     /db_xref="GI:4417267"
                     /translation="MILSFRLFAWECHVACGEVASERVILSTSTSLKPNAKRKKNSKS
                     KKLKQRILRDHESKKTMRKMKHLSSIANDVVQRCSQELETTIDDLVKEFECQWKPGST
                     GGTYSKKFVEFCNSKVTSRVCENILERIKDGSFTRLTFDMMLAWQQPDADDNESYKEA
                     VGKESEDKRIQATLSPEQDDISLFYSDMMPLLVDHEPSVGEDAFVYLGSIIPLPVDII
                     NGRYTFETLTAPTGHQLHFPAYDMFVKEIHKCMKHLQKQSTPKGIELADDEIILHVEG
                     TMASQRVIRHIKETSWPGRLTLTNYALYFEAAGIINYEDAIKIDLSKDNEKSTKPMST
                     GPLGAPLFDKAIVYESPDFEEGIVIEFPEMTSSTRRDHWLMLVKEITLMHKFLRKFNV
                     ESPLQSWEIHSRTILGIIRLHAAREMLRISPPDPKNFLIFSLFEEVPKGDYVLEELAE
                     ISLKIGTTRNPCSASSILRNMNMDQLGDMIKEEGEDICKEKVVKVTDKEEMLASLESA
                     VNQSREEGKVIEKARATTAELEEEGISESVAVLMKLNLCYREWVGKAIAACLIWVVAK
                     MAQARNKMVHTKSEDAVTVSTESDQTVTESIVSAQYGLIRLHQLMQHVNVTILKLRSL
                     YTSKASKHASMVMALMLVLASFFAVVPFKLFIIFGIVYCFVMTSSVGTYMSNDQSNRR
                     MKEWWDSIPIVPVRVRNASSK"
     repeat_region   complement(8503. .8530)
                     /rpt_family="AT_rich"
     repeat_region   complement(8755. .8807)
                     /rpt_family="(TA)n"
     repeat_region   complement(10076. .10140)
                     /rpt_family="AT_rich"
     repeat_region   complement(10600. .10646)
                     /rpt_family="AT_rich"
     gene            10941. .12546
                     /gene="At2g21730"
                     /note="synonym: F7D8.5"
     mRNA            join(10941. .11056,11162. .11275,11514. .12187,12284. .>12546)
                     /gene="At2g21730"
     CDS             join(10977. .11056,11162. .11275,11514. .12187,12284. .12546)
                     /gene="At2g21730"
                     /codon_start=1
                     /product="putative cinnamyl-alcohol dehydrogenase"
                     /protein_id="AAD20393.1"
                     /db_xref="GI:4417268"
                     /translation="MVDQNKAFGWAANDESGVLSPFHPSRRENGENDVTVKILFCGVC
                     HSDLHTIKNHWGFSRYPIIPGHEIVGIATKVGKNVTKFKEGDRVGVGVIIGSCQSCES
                     CNQDLENYCPKVVFTYNSRSSDGTSRNQGGYSDVIVVDHRFVLSIPDGLPSDSGAPLL
                     CAGITVYSPMKYYGMTKESGKRLGVNGLGGLGHIAVKIGKAFGLRVTVISRSSEKERE
                     AIDRLGADSFLVTTDSQKMKEAVGTMDFIIDTVSAEHALLPLFSLLKVNGKLVALGLP
                     EKPLDLPIFSLVLGRKMVGGSQIGGMKETQEMLEFCAKHKIVSDIELIKMSDINSAMD
                     RLAKSDVRYRFVIDVANSLLPESSAEILTEQVDHGVSITSRF"
     gene            complement(12909. .13286)
                     /gene="At2g21740"
                     /note="synonym: F7D8.6; predicted by genscan"
     mRNA            complement(<12909. .>13286)
                     /gene="At2g21740"
     CDS             complement(12909. .13286)
                     /gene="At2g21740"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAD20394.1"
                     /db_xref="GI:4417269"
                     /translation="MASNTSFLFATIAILLVLNISGRTLPETEDSTNIAARLNGGGLM
                     ECWNALYELKSCTNEIVLFFLNGETKLGVDCCQAVEVITTDCWPAMLTSLGFTSDETN
                     VLRGFCQSPNSGGSSPAPSSVKL"
     repeat_region   complement(13306. .13330)
                     /rpt_family="(CAAAA)n"
     gene            complement(14290. .14667)
                     /gene="At2g21750"
                     /note="synonym: F7D8.7; predicted by genscan"
     mRNA            complement(<14290. .>14667)
                     /gene="At2g21750"
```

```
     CDS             complement(14290. .14667)
                     /gene="At2g21750"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAD20395.1"
                     /db_xref="GI:4417270"
                     /translation="MASNTSFLFVTVTLLLVLNVSSRALPPVADSTNIAARLTGGGLM
                     QCWDALYELKSCTNEIVLFFLNGETKLGYGCCNAVDVITTDCWPAMLTSLGFTLEETN
                     VLRGFCQSPNSGGSSPALSPVKL"
     gene            14957. .15028
                     /gene="At2g21760"
     tRNA            14957. .15028
                     /gene="At2g21760"
                     /product="tRNA-Leu"
     repeat_region   complement(15276. .15302)
                     /rpt_family="(CAA)n"
     gene            15760. .20459
                     /gene="At2g21770"
                     /note="synonym: F7D8.9"
     mRNA            join(<15760. .15840,16269. .16464,16541. .16743,16827. .16929,
                     17017. .17209,17293. .17905,18005. .18142,18220. .18345,
                     18427. .18639,18725. .18992,19086. .19282,19367. .19717,
                     19834. .20459)
                     /gene="At2g21770"
     CDS             join(15760. .15840,16269. .16464,16541. .16743,16827. .16929,
                     17017. .17209,17293. .17905,18005. .18142,18220. .18345,
                     18427. .18639,18725. .18992,19086. .19282,19367. .19717,
                     19834. .20418)
                     /gene="At2g21770"
                     /codon_start=1
                     /product="putative cellulose synthase catalytic subunit"
                     /protein_id="AAD20396.1"
                     /db_xref="GI:4417271"
                     /translation="MNTGGRLIAGSHNRNEFVLINADDTARIRSAEELSGQTCKICRD

 Query Match              1.1%;  Score 39;  DB 8;  Length 107848;
 Best Local Similarity   100.0%;  Pred. No. 2.4e-09;
 Matches   39;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

/       1586 GGATTTGATCACCATAAGAAGGCTGGAGCTATGAATTCC 1624
             |||||||||||||||||||||||||||||||||||||||
)      18304 GGATTTGATCACCATAAGAAGGCTGGAGCTATGAATTCC 18342


:SULT 8
'008832
)CUS        BT008832               1041 bp    mRNA    linear   PLN 13-JUN-2003
:FINITION   Arabidopsis thaliana At5g09870 mRNA, complete cds.
:CESSION    BT008832
:RSION      BT008832.1  GI:31711829
:YWORDS     FLI_CDNA.
)URCE       Arabidopsis thaliana (thale cress)
 ORGANISM   Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
:FERENCE    1  (bases 1 to 1041)
 AUTHORS    Kim,C.J., Chen,H., Cheuk,R., Shinn,P., Bowser,L., Carninci,P.,
            Dale,J.M., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A.,
            Karlin-Neumann,G., Kawai,J., Lam,B., Lin,J., Miranda,M.,
            Narusaka,M., Nguyen,M., Onodera,C.S., Palm,C.J., Quach,H.L.,
            Sakurai,T., Satou,M., Seki,M., Southwick,A., Toriumi,M., Wong,C.,
            Wu,H.C., Yamada,K., Yu,G., Shinozaki,K., Davis,R.W., Theologis,A.
            and Ecker,J.R.
 TITLE      Arabidopsis ORF clones
 JOURNAL    Unpublished
:FERENCE    2  (bases 1 to 1041)
 AUTHORS    Kim,C.J., Chen,H., Cheuk,R., Shinn,P., Bowser,L., Carninci,P.,
            Dale,J.M., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A.,
            Karlin-Neumann,G., Kawai,J., Lam,B., Lin,J., Miranda,M.,
            Narusaka,M., Nguyen,M., Onodera,C.S., Palm,C.J., Quach,H.L.,
            Sakurai,T., Satou,M., Seki,M., Southwick,A., Toriumi,M., Wong,C.,
            Wu,H.C., Yamada,K., Yu,G., Shinozaki,K., Davis,R.W., Theologis,A.
            and Ecker,J.R.
 TITLE      Direct Submission
 JOURNAL    Submitted (13-JUN-2003) Salk Institute Genomic Analysis Laboratory
            (SIGnAL), Plant Biology Laboratory, The Salk Institute for
            Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037,
            USA
 COMMENT    RIKEN Genomic Sciences Center (GSC) members carried out the
            collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN
            Arabidopsis Full Length cDNA') : Seki,M., Narusaka,M., Ishida,J.,
            Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J.,
            Hayashizaki,Y. and Shinozaki,K.

            The Salk, Stanford, PGEC (SSP) Consortium members constructed and
            sequenced the pDNI (ORF) clones using the RAFL cDNAs: Kim,C.J.,
            Chen,H., Cheuk,R., Shinn,P., Bowser,L., Chan, M.M., Chang,C.M.,
            Dale,J.M., Jones,T., Karlin-Neumann,G., Lam,B., Lee,J.M., Lin,J.,
            Miranda,M., Nguyen,M., Onodera, C.S., Palm,C.J., Quach,H.L.,
            Southwick,A., Tang,C.C., Toriumi,M., Wong,C., Wu,H.C., Yamada,K.,
            Yu,G., Davis,R.W., Theologis,A., and Ecker,J.R.

            Kim,C.J. (SSP/Salk) and Seki,M. (RIKEN GSC) contributed equally to
            this work. Shinozaki,K. (RIKEN GSC) and Ecker,J.R. (SSP/Salk)
            contributed equally to this work as PIs
 FEATURES             Location/Qualifiers
     source           1. .1041
                      /organism="Arabidopsis thaliana"
                      /mol_type="mRNA"
                      /db_xref="taxon:3702"
                      /chromosome="5"
                      /clone="U24437"
                      /note="This clone is in pDNI 51
                      ecotype: Columbia"
     CDS              1. .1041
                      /note="cellulose synthase catalytic subunit"
                      /codon_start=1
                      /product="At5g09870"
                      /protein_id="AAP68271.1"
                      /db_xref="GI:31711830"
                      /translation="MENGGLARNASPASLLREAIQVISCGYEDKTEWGKEIGWIYGSV
                      TEDILTGFKMHSHGWRSVYCTPKIPAFKGSAPINLSDRLHQVLRWALGSVEIFLSRHC
                      PIWYGYGGGLKWLERLSYINSVVYPWTSIPLLVYCSLPAICLLTGKFIVPEISNYAST
                      LFMALFGSIAVTGILEMQWGKVGIDDWWRNEQFWVIGGVSAHLFALFQGLLKVLAGVE
                      TNFTVTSKAADDGEFSELYIFKWTSLLIPPTTLLIINVIGVVVGISDAISNGYDSWGP
                      LFGRLFFAFWVILHLYPFLKGLLGKQDRMPTIILVWSILLASILTLLWVRVNPFVAKG
                      GPILEICGLDCL"
 BASE COUNT       238 a    226 c    247 g    330 t
 ORIGIN

   Query Match              1.1%;  Score 38;  DB 8;  Length 1041;
   Best Local Similarity   100.0%;  Pred. No. 9.7e-09;
   Matches   38;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

 Qy       2753 GGGATCGATGATTGTGGAGAAACGAACAGTTTTGGGT 2790
               ||||.|||.||| ||||||||| |||.|||||
 Db        547 GGGATCGATGATTGTGGAGAAACGAACAGTTTTGGGT 584


 RESULT 9
 AY136423
 LOCUS       AY136423               1490 bp    mRNA    linear   PLN 27-AUG 2002
 DEFINITION  Arabidopsis thaliana cellulose synthase catalytic subunit
             (At5g09870) mRNA, complete cds.
 ACCESSION   AY136423
 VERSION     AY136423.1  GI:22531169
 KEYWORDS    FLI_CDNA.
 SOURCE      Arabidopsis thaliana (thale cress)
   ORGANISM  Arabidopsis thaliana
             Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
             Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
             rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 REFERENCE   1  (bases 1 to 1490)
   AUTHORS   Southwick,A., Nguyen,M., Tripp,M., Palm,C.J., Jones,T., Wu,T.,
             Carninci,P., Chen,H., Cheuk,R., Chan,M.M., Chang,C.H., Dale,J.M.,
```

```
            Deng,J.M., Hayashizaki,Y., Hsuan,V.W., Lee,J.M., Ishida,J.,
            Kamiya,A., Kawai,J., Kim,C.J., Narusaka,M., Quach,H.L., Sakurai,T.,
            Satou,M., Seki,M., Shinn,P., Tang,C.C., Toroumi,M., Wallender,E.K.,
            Wong,C., Wu,H.C., Yamada,K., Yu,G., Yuan,S., Shinozaki,K.,
            Ecker,J., Theologis,A. and Davis,R.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JUL-2002) DNA Sequencing and Technology Center,
            Stanford University, 855 California Avenue, Palo Alto, CA 94304,
            USA
COMMENT     e-mail for correspondence: arabs@sequence.stanford.edu

            RIKEN Genomic Sciences Center (GSC) members carried out the
            collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN
            Arabidopsis  Full-Length cDNA'): Seki,M., Narusaka,M., Ishida,J.,
            Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J.,
            Hayashizaki,Y. and Shinozaki,K.

            The Salk, Stanford, PGEC (SSP) Consortium members carried out the
            sequencing and annotation of the RAFL cDNAs: Southwick,A.,
            Nguyen,M., Tripp,M., Palm,C.J., Jones,T., Wu,T., Chen,H., Cheuk,R.,
            Chan,M.M., Chang,C.H., Dale,J.M., Deng,J.M., Hsuan,V.W., Lee,J.M.,
            Kim,C.J., Quach,H.L., Shinn,P., Tang,C.C., Toroumi,M.,
            Wallender,E.K., Wong,C., Wu,H.C., Yamada,K., Yu,G., Yuan,S.,
            Ecker,J., Theologis,A. and Davis,R.W.

            Southwick,A. (SSP/Stanford) and Seki,M. (RIKEN GSC) contributed
            equally to this work.  Shinozaki,K. (RIKEN GSC) and Davis,R.W.
            (SSP/Stanford) contributed equally to this work as PIs.
FEATURES             Location/Qualifiers
     source          1. .1490
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
                     /chromosome="5"
                     /clone="RAFL09-45-N15 (R24437)"
                     /note="This clone is in pBluescript"
                     ecotype: Columbia"
     gene            1. .1490
                     /gene="At5g09870"
     CDS             160. .1200
                     /gene="At5g09870"
                     /codon_start=1
                     /product="cellulose synthase catalytic subunit"
                     /protein_id="AAM97089.1"
                     /db_xref="GI:22531170"
                     /translation="MENGGLARNASPASLLREAIQVISCGYEDKTEWGKEIGWIYGSV
                     TEDILTGFKMHSHGWRSVYCTPKRPAFKGSAPINLSDRLHQVLRWALGSVEIFLSRHC
                     PIWYGYGGGLKWLERLSYINSVVYPWTSIPLLVYCILPAICLLTGKFIVPEISNYASI
                     LFMALFGSIAVTGILEMQWSGVGIDEWWRNEQFWVIGGVSAHLFAVFQGLLKVLAGVE
                     TNFTVTSKAADDGEFSELYIFKWTSLLIPPITLLMVNIVGIVSDAISNGYDSWGP
                     LFGRLFFAFWVILHLYPFLKGLMGRQNRTPTIVLVWSGLIAVISLLWVRVNPFVAKG
                     GPILEICGLDCL"
BASE COUNT       487 a    281 c    364 g    468 t
ORIGIN

  Query Match             1.1%;  Score 38;  DB 8;  Length 1490;
  Best Local Similarity   100.0%;  Pred. No. 9.6e-09;
  Matches   38;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2753 GGGATCGATGATTGGTGGAGAAACGAACAGTTTTGGGT 2790
              ||||||||||||||||||||||||||||||||||||||
Db        706 GGGATCGATGATTGGTGGAGAAACGAACAGTTTTGGGT 743


RESULT 10
AC125473/c
LOCUS       AC125473              103966 bp    DNA     linear   PLN 22-FEB-2003
DEFINITION  Medicago truncatula clone mth2-8j13, complete sequence.
ACCESSION   AC125473
VERSION     AC125473.12  GI:28467165
KEYWORDS    HTG.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 103966)
  AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
  TITLE     Medicago truncatula BAC Clone mth2-8j13
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 103966)
  AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JUN-2002) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   3  (bases 1 to 103966)
  AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-OCT-2002) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   4  (bases 1 to 103966)
  AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-FEB-2003) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
COMMENT     On Feb 22, 2003 this sequence version replaced gi:24111008.
            -------------- Genome Center
            Center: Department Of Chemistry And Biochemistry
            The University Of Oklahoma
            Center code:UOKNOR
            ----------------------.
FEATURES             Location/Qualifiers
     source          1. .103966
                     /organism="Medicago truncatula"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3880"
                     /clone="mth2-8j13"
                     /clone_lib="Medicago truncatula BAC library H2"
BASE COUNT    35271 a   16915 c   16766 g   35014 t
ORIGIN

  Query Match             1.1%;  Score 38;  DB 8;  Length 103966;
  Best Local Similarity   100.0%;  Pred. No. 8.8e-09;
  Matches   38;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        170 GAATGTGCATTCCCTGTGTGTAGACCTTGCTATGAGTA 207
              ||||||||||||||||||||||||||||||||||||||
Db      12573 GAATGTGCATTCCCTGTGTGTAGACCTTGCTATGAGTA 12536


RESULT 11
AC126014
LOCUS       AC126014              113199 bp    DNA     linear   PLN 24-APR-2003
DEFINITION  Medicago truncatula clone mth2-18j5, complete sequence.
ACCESSION   AC126014
VERSION     AC126014.5  GI:24414363
KEYWORDS    HTG.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 113199)
  AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
  TITLE     Medicago truncatula BAC Clone mth2-18j5
  JOURNAL   Unpublished
```

```
EFERENCE   2  (bases 1 to 113199)
 AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
           Cook,D., Kim,D. and Roe,B.A.
 TITLE     Direct Submission
 JOURNAL   Submitted (02-JUL-2002) Department Of Chemistry And Biochemistry,
           The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
           OK 73019, USA
EFERENCE   3  (bases 1 to 113199)
 AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
           Cook,D., Kim,D. and Roe,B.A.
 TITLE     Direct Submission
 JOURNAL   Submitted (09-JUL-2002) Department Of Chemistry And Biochemistry,
           The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
           OK 73019, USA
EFERENCE   4  (bases 1 to 113199)
 AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
           Cook,D., Kim,D. and Roe,B.A.
 TITLE     Direct Submission
 JOURNAL   Submitted (19-JUL-2002) Department Of Chemistry And Biochemistry,
           The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
           OK 73019, USA
EFERENCE   5  (bases 1 to 113199)
 AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
           Cook,D., Kim,D. and Roe,B.A.
 TITLE     Direct Submission
 JOURNAL   Submitted (06-AUG-2002) Department Of Chemistry And Biochemistry,
           The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
           OK 73019, USA
EFERENCE   6  (bases 1 to 113199)
 AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
           Cook,D., Kim,D. and Roe,B.A.
 TITLE     Direct Submission
 JOURNAL   Submitted (08-AUG-2002) Department Of Chemistry And Biochemistry,
           The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
           OK 73019, USA
EFERENCE   7  (bases 1 to 113199)
 AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
           Cook,D., Kim,D. and Roe,B.A.
 TITLE     Direct Submission
 JOURNAL   Submitted (09-AUG-2002) Department Of Chemistry And Biochemistry,
           The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
           OK 73019, USA
EFERENCE   8  (bases 1 to 113199)
 AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
           Cook,D., Kim,D. and Roe,B.A.
 TITLE     Direct Submission
 JOURNAL   Submitted (01-SEP-2002) Department Of Chemistry And Biochemistry,
           The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
           OK 73019, USA
EFERENCE   9  (bases 1 to 113199)
 AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
           Cook,D., Kim,D. and Roe,B.A.
 TITLE     Direct Submission
 JOURNAL   Submitted (26-OCT-2002) Department Of Chemistry And Biochemistry,
           The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
           OK 73019, USA
EFERENCE   10 (bases 1 to 113199)
 AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
           Cook,D., Kim,D. and Roe,B.A.
 TITLE     Direct Submission
 JOURNAL   Submitted (06-DEC-2002) Department Of Chemistry And Biochemistry,
           The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
           OK 73019, USA
EFERENCE   11 (bases 1 to 113199)
 AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
           Cook,D., Kim,D. and Roe,B.A.
 TITLE     Direct Submission
 JOURNAL   Submitted (24-APR-2003) Department Of Chemistry And Biochemistry,
           The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
           OK 73019, USA
OMMENT     On Oct 26, 2002 this sequence version replaced gi:22138507.
           ---------- Genome Center
           Center: Department Of Chemistry And Biochemistry
           The University Of Oklahoma
           Center code:UOKC_B
           ---------- ---- ----.
FEATURES             Location/Qualifiers
     source          1..113199
                     /organism="Medicago truncatula"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3880"
                     /clone="mth2-18j5"
                     /clone_lib="Medicago truncatula BAC library H1"
BASE COUNT    37854 a   18124 c   18388 g   38833 t
ORIGIN

  Query Match                1%;  Score 38;  DB 8;  Length 113199;
  Best Local Similarity   1.0%;  Pred. No. 8.8e-09;
  Matches   38;  Conservati      0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        120 GAATGGTCAATTC  GTTGCAAA  TTTATAGGTA 202
              ||||| |||  |||   ||   | |   ||| ||||
Db      87265 GAATGGTCAATTC  GTTGCAAA  TTTATAGGTA 87202


RESULT 12
AF030052
LOCUS       AF030052                1838 bp    mRNA    linear   PLN 05-FEB-1998
DEFINITION  Oryza sativa subsp. japonica RSW1 like cellulose synthase catalytic
            subunit mRNA, partial cds
ACCESSION   AF030052
VERSION     AF030052.1  GI:2781433
KEYWORDS    .
SOURCE      Oryza sativa (japonica cultivar group)
  ORGANISM  Oryza sativa (japonica cultivar group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza
REFERENCE   1  (bases 1 to 1838)
  AUTHORS   Arioli,T., Peng,L., Betzner,A.S., Burn,J., Wittke,W., Herth,W.,
            Camilleri,C., Hofte,H., Plazinski,J., Birch,R., Cork,A., Glover,J.,
            Redmond,J. and Williamson,R.E.
  TITLE     Molecular analysis of cellulose biosynthesis in Arabidopsis
  JOURNAL   Science 279 (5351), 717-720 (1998)
  MEDLINE   98111412
   PUBMED   9445479
REFERENCE   2  (bases 1 to 1838)
  AUTHORS   Burn,J.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-OCT-1997) Plant Science Centre, Australian National
            University, Acton, Canberra, ACT 0200, Australia
FEATURES             Location/Qualifiers
     source          1..1838
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="Nipponbare"
                     /sub_species="japonica"
                     /db_xref="taxon:39947"
     CDS             90..2501838
                     /codon_start=1
                     /product="RSW1-like cellulose synthase catalytic subunit"
                     /protein_id="AAC19333.1"
                     /db_xref="GI:2781433"
                     /translation="MAANAGMVAGSRNRNEFVMIRPDGDAPPPAKPGKSVNGQVCQIC
                     GDTVGVSATGDVFVACNECAFPVCRPCYEYERKEGNQCCPQCKTRYKRHKGSPRVQGD
                     EEEEDVDDLDNEFNYKHGNGKGPEWQIQRQGEDVDLSSSSSRHEQHRIPRLTSGQQISG
                     EIPDASHGRHSIRSGTSSYVDPSVPVPVRIVDPSKDLNSYGINSVDWQERVASWRNKQ
                     DKNMMQVANKYPEARGGDMEGTGSNGEDIQMVDDARLPLSRIVPIPSNQLNLYRIVI
                     LRLIILMFFFQYRVTHPVRDAYGLWLVSVICEIWFALSWLLDQFPKWYPINRETYLDR
                     LALRYDREGEPSQLAPIDVFVSTVDPLKEPPLITANTVLSILAVDYPVDKVSCYVSDD
                     GSAMLTFEALSETAEFARKWVPFCKKHNIEPRAPEFYFAQKIDYLKDKIQPSFVKERR
                     AMKREYEEFKVRINALVAKAQKVPEEGWTMADGTAWPGNNPRDHPGMIQVFLGHSGGL
                     DTDGNELPRLVYVSREKRPGFQHHKKAGAMNALIRVSAVLTNGAYLLNVDCDHYFNNS
                     KALREAMCFMMDPALGR"
BASE COUNT      472 a    414 c    507 g    445 t
ORIGIN
```

```
  Query Match             1.0%;  Score 34;  DB 8;  Length 1838;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-06;
  Matches   34;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1692 TCAACAATAGCAAAGCTCTTAGAGAAGCAATGTG 1725
             ||||||||||||||||||||||||||||||||||
Db       1777 TCAACAATAGCAAAGCTCTTAGAGAAGCAATGTG 1810


RESULT 13
AX652952
LOCUS       AX652952                3222 bp    DNA     linear   PAT 22-MAR-2003
DEFINITION  Sequence 2822 from Patent WO03000898.
ACCESSION   AX652952
VERSION     AX652952.1  GI:29155766
KEYWORDS    .
SOURCE      Oryza sativa
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza
REFERENCE   1
  AUTHORS   Chang,H.S., Chen,W., Cooper,B., Glazebrook,J., Goff,S.A., Hou,Y.M.,
            Katagiri,F., Quan,S., Tao,Y., Whitham,S., Xie,Z., Zhu,T. and Zou,G.
  TITLE     Plant genes involved in defense against pathogens
  JOURNAL   Patent: WO 03000898-A 2822 03-JAN 2003;
            Syngenta Participations AG (CH)
FEATURES             Location/Qualifiers
     source          1. .3222
                     /organism="Oryza sativa"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4530"
BASE COUNT      805 a    711 c    809 g    901 t
ORIGIN

  Query Match             0.9%;  Score 32;  DB 6;  Length 3222;
  Best Local Similarity   100.0%;  Pred. No. 2.2e-05;
  Matches   32;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1679 TGTGATCACTACATCAACAATAGCAAAGCTCT 1710
             ||||||||||||||||||||||||||||||||
Db       1666 TGTGATCACTACATCAACAATAGCAAAGCTCT 1697


RESULT 14
AX505864
LOCUS       AX505864                3255 bp    DNA     linear   PAT 27-SEP-2002
DEFINITION  Sequence 559 from Patent WO0216655.
ACCESSION   AX505864
VERSION     AX505864.1  GI:23787131
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Harper,J.F., Kreps,J., Wang,X. and Zhu,T.
  TITLE     Stress-regulated genes of plants, transgenic plants containing
            same, and methods of use
  JOURNAL   Patent: WO 0216655-A 559 28-FEB 2002;
            The Scripps Research Institute (US) ; Syngenta Participations AG
            (CH)
FEATURES             Location/Qualifiers
     source          1. .3255
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3702"
BASE COUNT      884 a    621 c    798 g    952 t
ORIGIN

  Query Match             0.9%;  Score 32;  DB 6;  Length 3255;
  Best Local Similarity   100.0%;  Pred. No. 2.2e-05;
  Matches   32;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       419 TTGGATTCAGCTCCACCTGGCTCTCAGATTCC 450
             ||||||||||||||||||||||||||||||||
Db       415 TTGGATTCAGCTCCACCTGGCTCTCAGATTCC 446


RESULT 15
AF027173
LOCUS       AF027173                3771 bp    mRNA    linear   PLN 07-FEB-1998
DEFINITION  Arabidopsis thaliana cellulose synthase catalytic subunit (Ath-A)
            mRNA, complete cds.
ACCESSION   AF027173
VERSION     AF027173.1  GI:2827140
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 3771)
  AUTHORS   Arioli,T., Peng,L., Betzner,A.S., Burn,J., Wittke,W., Herth,W.,
            Camilleri,C., Hofte,H., Plazinski,J., Birch,R., Cork,A., Glover,J.,
            Redmond,J. and Williamson,R.E.
  TITLE     Molecular analysis of cellulose biosynthesis in Arabidopsis
  JOURNAL   Science 279 (5351), 717-720 (1998)
  MEDLINE   98111412
   PUBMED   9445479
REFERENCE   2  (bases 1 to 3771)
  AUTHORS   Arioli,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-SEP-1997) Plant Science Centre, Australian National
            University, Acton, Canberra, ACT 200, Australia
FEATURES             Location/Qualifiers
     source          1. .3771
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /cultivar="Columbia"
                     /db_xref="taxon:3702"
     gene            1. .3771
                     /gene="Ath-A"
     CDS             217. .3471
                     /gene="Ath-A"
                     /note="RSW1-like"
                     /codon_start=1
                     /product="cellulose synthase catalytic subunit"
                     /protein_id="AAC39335.1"
                     /db_xref="GI:2827141"
                     /translation="MNTGGRLIAGSHNRNEFVLINADESARIRSVQELSGQTCQICGD
                     EIELTVSSELFVACNECAFPVCRPCYEYERREGNQACPQCKTRYKRIKGSPRVDGDDE
                     EEEDIDDLEYEFDHGMDPEHAAEAALSSRLNTGRGGLDSAPFGSQIPLLTYCDEDADM
                     YSDRHALIVPPSTGYGNRVYPAPFTDSSAPPQARSMVPQKDIAEYGYGSVAWKDRMEV
                     WKRRQGEKLQVIKHEGGNNGRGSNDDDELDDPDMPMMDEGRQPLSRKLPIRSSRINPY
                     RMLILCRLAILGLFFHYRILHPVNDAYGLWLTSVICEIWFAVSWILDQFPKWYPIERE
                     TYLDRLSLRYEKEGKPSGLAPVDVFVSTVDPLKEPPLITANTVLSILAVDYPVDKVAC
                     YVSDDGAAMLTFEALSDTAEFARKWVPFCKKFNIEPRAPEWYFSQKMDYLKNKVHPAF
                     VRERRAMKRDYEEFKVKINALVATAQKVPEEGWTMQDGTPWPGNNVRDHPGMIQVFLG
                     HSGVRDTDGNELPRLVYVSREKRPGFDHHKKAGAMNSLIRVSAVLSNAPYLLNVDCDH
                     YINNSKAIRESMCFMMDPQSGKKVCYVQFPQRFDGIDRHDRYSNRNVVFFDINMKGLD
                     GIQGPIYVGTGCVFRRQALYGFDAPKKKKPPGKTCNCWPKWCCLCCGLRKKSKTKAKD
                     KKTNTKETSKQIHALENVDEGVIVPVSNVEKRSEATQLKLEKKFGQSPVFVASAVLQN
                     GGVPRNASPACLLREAIQVISCGYEDKTEWGKEIGWIYGSVTEDILTGFKMHCHGWRS
                     VYCMPKRAAFKGSAPINLSDRLHQVLRWALGSVEIFLSRHCPIWYGYGGGLKWLERFS
                     YINSVVYPWTSLFLIVYCSLPAVCLLTGKFIVPEISNYAGILFMLMFISIAVTGILEM
                     QWGGVGIDDWWRNEQFWVIGGASSHLFALFQGLLKVLAGVNTNFTVTSKAADDGAFSE
                     LYIFKWTTLLIPPTTLLIINIIGVIVGVSDAISNGYDSWGPLFGRLFFALWVIVHLYP
                     FLKGMLGKQDKMPTIIVVWSILLASILTLLWVRVNPFVAKGGPVLEICGLNCGN"
BASE COUNT     1037 a    699 c    907 g   1128 t
ORIGIN

  Query Match             0.9%;  Score 32;  DB 8;  Length 3771;
  Best Local Similarity   100.0%;  Pred. No. 2.2e-05;
```

Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
        419 TTGGATTCAGCTCCACCTGGCTCTCAGATTCC 450
            ||||||||||||||||||||||||||||||||
        631 TTGGATTCAGCTCCACCTGGCTCTCAGATTCC 662
```

arch completed: October 31, 2003, 05:47:08
ob time : 12270.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 30, 2003, 16:25:58 ; Search time 837.27 Seconds
(without alignments)
11103.788 Million cell updates/sec

Title: US-09-980-043-5
Perfect score: 3444
Sequence: 1 gcggcggcggttaatcgccg.....gtgttgtgatagcggccgc 3444

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2552756 seqs, 1349719017 residues

Word size : 0

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3444	100.0	3444	22	AAC83798	Arabidopsis thalia
2	2322	67.4	3255	24	ABZ14725	Arabidopsis thalia
c 3	361	10.5	447	24	ABL93908	Arabidopsis thalia
c 4	222	6.4	342	25	ABX61330	Arabidopsis thalia
5	32	0.9	2890	21	AAZ58270	Soybean cellulose
6	32	0.9	3255	24	ABZ12754	Arabidopsis thalia
7	32	0.9	3828	19	AAV06566	Arabidopsis cellul
8	32	0.9	5009	19	AAV06564	Arabidopsis cellul
9	29	0.8	3207	19	AAV08372	Cellulose synthase
10	29	0.8	3328	19	AAV34432	Cotton cellulose s
11	29	0.8	3704	21	AAZ99533	DNA encoding a mai
12	27	0.8	3517	21	AAZ58268	Soybean cellulose
13	26	0.8	3725	21	AAZ99500	DNA encoding a mai
14	26	0.8	3725	21	AAZ99506	DNA encoding a mai
15	26	0.8	3725	21	AAZ99521	DNA encoding a mai
16	26	0.8	3753	21	AAZ99515	DNA encoding a mai
17	26	0.8	3753	21	AAZ99530	DNA encoding a mai
18	26	0.8	3780	21	AAZ99497	DNA encoding a mai
19	26	0.8	3786	21	AAZ58265	Corn cellulose syn
20	26	0.8	3799	24	AAS16458	Corn cDNA encoding
21	26	0.8	3813	21	AAZ99509	DNA encoding a mai
22	26	0.8	3813	21	AAZ99524	DNA encoding a mai
23	24	0.7	7234	22	AAC83247	Cellulose synthase
24	23	0.7	360	24	ABQ85650	Arabidopsis thalia
25	23	0.7	950	21	AAC50201	Arabidopsis thalia
26	23	0.7	951	21	AAC33457	Arabidopsis thalia
27	23	0.7	3232	22	AAC65448	Populus tremuloide
28	23	0.7	3936	21	AAZ58266	Corn cellulose syn
29	23	0.7	3969	21	AAZ99503	DNA encoding a mai
30	23	0.7	3969	21	AAZ99518	DNA encoding a mai
31	23	0.7	27179	22	AAS34556	Human DNA for a no
32	22	0.6	353	22	AAL25959	Human breast cance
33	22	0.6	366	22	AAL19124	Human breast cance
34	22	0.6	427	22	AAL08214	Human breast cance
35	22	0.6	439	25	ABX50372	Bovine EST associa
36	22	0.6	440	25	ABX50557	Bovine EST associa
37	22	0.6	818	21	AAF09881	Fusarium venenatum
38	22	0.6	1086	19	AAV08374	Cellulose synthase
39	22	0.6	2316	22	AAS34805	cDNA encoding nove
40	22	0.6	2471	24	ABQ54703	Human ovarian anti
41	22	0.6	2487	21	AAF15969	Human prostate can
42	22	0.6	2517	20	AAZ28269	Rat neuronal immed
c 43	22	0.6	4667	23	ABL05017	Drosophila melanog
c 44	22	0.6	7465	23	ABL05016	Drosophila melanog
45	22	0.6	11489	22	AAK86324	Human immune/haema

ALIGNMENTS

RESULT 1
AAC83798
ID AAC83798 standard; cDNA; 3444 BP.
XX
AC AAC83798;
XX
DT 07-MAR-2001 (first entry)
XX
DE Arabidopsis thaliana cellulose synthase CelA cDNA.
XX
KW Arabidopsis thaliana; aspen; CelA; cellulose synthase; wood;
KW mechanical stress response element; MSRE; cellulose biosynthesis; ss.
XX
OS Arabidopsis thaliana.
XX
PN WO200071670-A2.
XX
PD 30-NOV-2000.
XX
PF 18-MAY-2000; 2000WO-US13637.
XX
PR 21-MAY-1999; 99US-0135280.
XX
PA (UNMT) UNIV MICHIGAN TECHNOLOGICAL.
XX
PI Chiang VL, Wu L, Chandrashekhar JP, Carraway DT;
XX
DR WPI; 2001-032025/04.
XX
PT Novel polynucleotide encoding cellulose synthase useful for enhancing
PT cellulose biosynthesis, modifying lignin biosynthesis, and for

accelerating growth of plants and for upregulating cellulose synthase levels -

Claim 8; Fig 7; 65pp; English.

The present sequence is given in a specification relating to polynucleotides encoding cellulose synthase enzymes from Populus tremuloides and Arabidopsis. The cellulose synthases are designated PtCelA and AraxCelA respectively. The polynucleotides may be introduced into plants where they are useful for altering the growth of a plant. Vectors comprising a cellulose synthase promoter, or a fragment of the promoter containing one or more mechanical stress response elements (MSRE), operably linked to a cellulose synthase gene are useful for causing stress-induced gene expression in a plant cell. A cDNA encoding a protein that binds to a positive MSRE may be delivered into the plant to increase expression of cellulose synthase. A cDNA in an antisense orientation may be delivered to reduce expression of cellulose synthase. The polynucleotides enable the development of transgenic tree varieties having increased cellulose content, decreased lignin content and therefore improved wood fibre characteristics. The cellulose synthase promoters are useful for synthesising greater quantities of high-crystalline cellulose in plant and preferably in trees, which permits an increase in proportion of cellulose in transgenic plants, greater strength of juvenile wood and fiber and acceleration of overall growth rate.
Note: This nucleotide sequence is given in the specification as SEQ ID NO: 5. SEQ ID NO: 5 is referred to in claim 8 but it is described as a polypeptide sequence.

Sequence 3444 BP; 903 A; 664 C; 860 G; 1017 T; 0 other;

Query Match 100.0%; Score 3444; DB 22; Length 3444;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3444; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GCGGCCGCGGTTAATCGCCGGTTCTCACAACAGGAATGAGTTTGTCCTCATTAATGCCGA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GCGGCCGCGGTTAATCGCCGGTTCTCACAACAGGAATGAGTTTGTCCTCATTAATGCCGA 60

Qy         61 TGAGAATGCCCGAATAAGATCAGTCCAAGAGCTGAGTGGACAGACATGTCAAATCTGCAG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TGAGAATGCCCGAATAAGATCAGTCCAAGAGCTGAGTGGACAGACATGTCAAATCTGCAG 120

Qy        121 AGATGAGATCGAATTGACTGTTGATGGAGAACCGTTTGTGGCATGTAACGAATGTGCATT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AGATGAGATCGAATTGACTGTTGATGGAGAACCGTTTGTGGCATGTAACGAATGTGCATT 180

Qy        181 CCCTGTGTGTAGACCTTGCTATGAGTACGAAAGACGAGAAGGCAATCAAGCTTGTCCACA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCCTGTGTGTAGACCTTGCTATGAGTACGAAAGACGAGAAGGCAATCAAGCTTGTCCACA 240

Qy        241 GTGCAAAACCCGTTTCAAACGTCTTAAAGGAAGTCCAAGAGTTGAAGGTGATGAAGAGGA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTGCAAAACCCGTTTCAAACGTCTTAAAGGAAGTCCAAGAGTTGAAGGTGATGAAGAGGA 300

Qy        301 AGATGACATTGATGATTTAGACAATGAGTTTGAGTATGGAAATAATGGGATTGGATTTGA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGATGACATTGATGATTTAGACAATGAGTTTGAGTATGGAAATAATGGGATTGGATTTGA 360

Qy        361 TCAGGTTTCTGAAGGTATGTCAATCTCTCGTCGCAACTCCGGTTTCCCACAATCTGATTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TCAGGTTTCTGAAGGTATGTCAATCTCTCGTCGCAACTCCGGTTTCCCACAATCTGATTT 420

Qy        421 GGATTCAGCTCCACCTGGCTCTCAGATTCCATTGCTGACTTACGGCGACGAGGACGTTGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGATTCAGCTCCACCTGGCTCTCAGATTCCATTGCTGACTTACGGCGACGAGGACGTTGA 480

Qy        481 GATTTCTTCTGATAGACATGCTCTTATTGTTCCTCCTTCACTTGGTGGTCATGGCAATAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GATTTCTTCTGATAGACATGCTCTTATTGTTCCTCCTTCACTTGGTGGTCATGGCAATAG 540

Qy        541 AGTTCATCCTGTTTCTCTTTCTGACCCGACCGTGGCTGCACATCGAAGGCTGATGGTACC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AGTTCATCCTGTTTCTCTTTCTGACCCGACCGTGGCTGCACATCGAAGGCTGATGGTACC 600

Qy        601 TCAGAAAGATCTTCAGGTTTATGGTTACTGAAGTGTCGCTTGGAAAGATCGGATGGAGGA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TCAGAAAGATCTTCAGGTTTATGGTTACTGAAGTGTCGCTTGGAAAGATCGGATGGAGGA 660

Qy        661 ATGGAAGAGAAAGTCAAATGAGAAATTTCAGGTTGTTAGGCATGAAGGAGATCCTGATTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ATGGAAGAGAAAGTCAAATGAGAAATTTCAGGTTGTTAGGCATGAAGGAGATCCTGATTT 720

Qy        721 TGAAGATGGTGATCAAGTTGATTTTCAATGATGGATGAGGGAAGGCAGCCATTGTCTAT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TGAAGATGGTGATCAAGTTGATTTTCAATGATGGATGAGGGAAGGCAGCCATTGTCTAT 780

Qy        781 GAAGATACCAATTTAATGGTAGAAGATAAATTGTTACGGATGTTAATTGTGCTACGTCT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAAGATACCAATTTAATGGTAGAAGATAAATTGTTACGGATGTTAATTGTGCTACGTCT 840

Qy        841 TGTGATTCTTGGTTTTCTTTCTTCATTATTCTTCACCCGTCAAAGATGCATATGC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TGTGATTCTTGGTTTTCTTTCTTCATTATTCTTCACCCGTCAAAGATGCATATGC 900

Qy        901 TTTGTGGCTTATTTTGTTATATGTGAGATATGGTTTGCTGTTTCATGGGTTCTTGATCA 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTTGTGGCTTATTTTGTTATATGTGAGATATGGTTTGCTGTTTCATGGGTTCTTGATCA 960

Qy        961 GTTCCCTAAAATGGTAGCCTATCGAGCGAGAAAGTACTTGGACCGACTCTCATTAAGATA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GTTCCCTAAAATGGTAGCCTATCGAGCGAGAAAGTACTTGGACCGACTCTCATTAAGATA 1020

Qy       1021 TGAGAAAGAAGGGAAACCGTCGGGACTATCCCTGTGGATGTATTTGTTAGTACAGTGGA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 TGAGAAAGAAGGGAAACCGTCGGGACTATCCCTGTGGATGTATTTGTTAGTACAGTGGA 1080

Qy       1081 TCCATTGAAAGAGCTTCCGCTTATTACTGCAAATACTGTCTTGTCTATTCTTGCTGTTGA 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 TCCATTGAAAGAGCTTCCGCTTATTACTGCAAATACTGTCTTGTCTATTCTTGCTGTTGA 1140

Qy       1141 TTATCCTGTCGATAAGGTTGCTTGTTACGTATCTGATGATGGTGCTGCTATGCTTACTTT 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 TTATCCTGTCGATAAGGTTGCTTGTTACGTATCTGATGATGGTGCTGCTATGCTTACTTT 1200

Qy       1201 CGAAGCTCTTTCTGAGACCGCTGAATTCGCAAGAAATGGGTTCCTTTCTGCAAGAAATA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 CGAAGCTCTTTCTGAGACCGCTGAATTCGCAAGAAATGGGTTCCTTTCTGCAAGAAATA 1260

Qy       1261 TTGTATTGAGCCTCGTGCTCCCGAATGGTATTTCTGCCATAAAATGGACTACTTGAAGAA 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 TTGTATTGAGCCTCGTGCTCCCGAATGGTATTTCTGCCATAAAATGGACTACTTGAAGAA 1320

Qy       1321 TAAAGTTCATCCCGCATTTGTTAGGGAGCGGCGAGCCATGAAGAGAGATTATGAAGAATT 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TAAAGTTCATCCCGCATTTGTTAGGGAGCGGCGAGCCATGAAGAGAGATTATGAAGAATT 1380

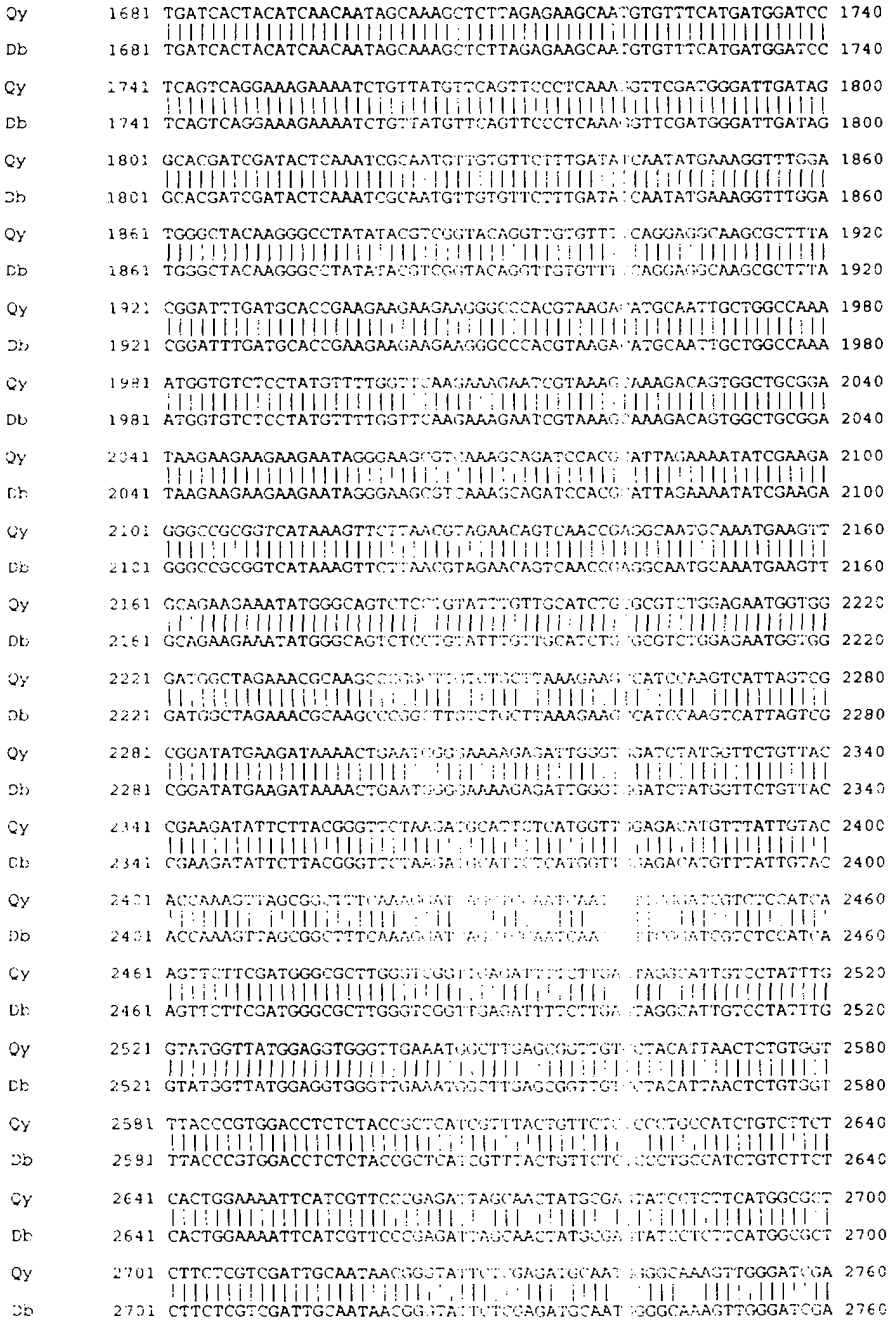
Qy       1381 CAAAGTAAAGATCAATGCTTTAGTAGCAACAGCACAGAAAGTGCCTGAGGATGGTTGGAC 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CAAAGTAAAGATCAATGCTTTAGTAGCAACAGCACAGAAAGTGCCTGAGGATGGTTGGAC 1440

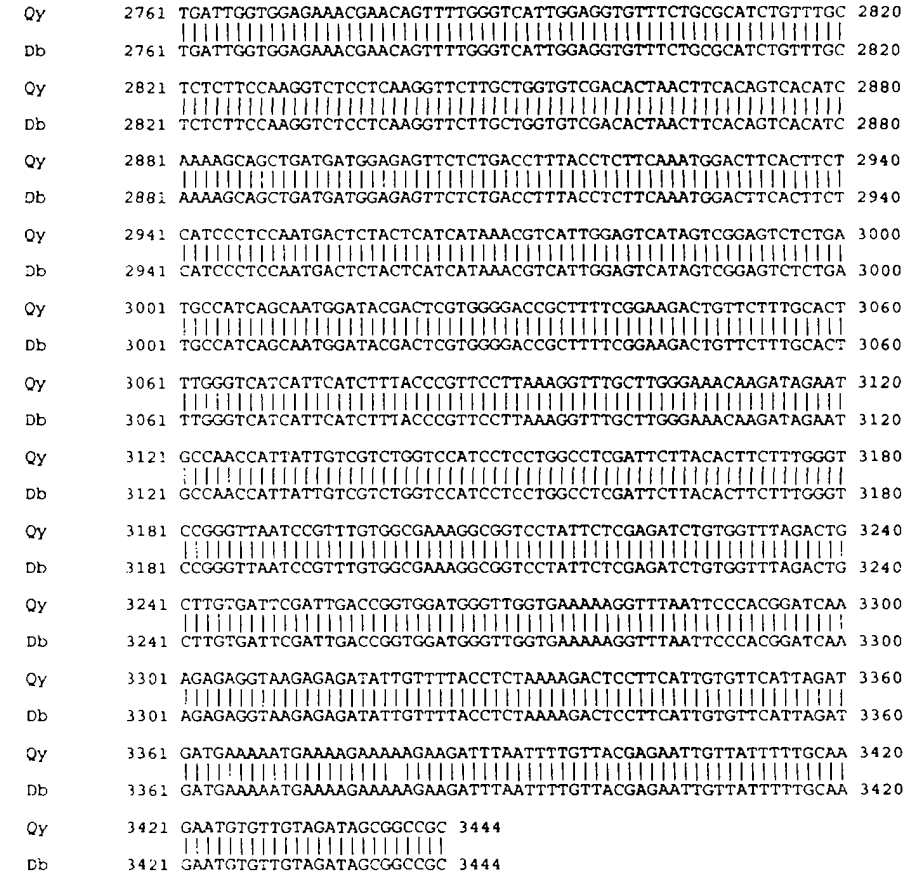
Qy       1441 TATGCAAGACGGTACACCTTGGCCCCGGTAATAGTGTGCGAGATCATCCTGGCATGATTCA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 TATGCAAGACGGTACACCTTGGCCCCGGTAATAGTGTGCGAGATCATCCTGGCATGATTCA 1500

Qy       1501 GGTCTTCCTTGGAAGTGACGGTGTTCGTGATGTCGAAAACAACGAGTTGCCTCGATTAGT 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 GGTCTTCCTTGGAAGTGACGGTGTTCGTGATGTCGAAAACAACGAGTTGCCTCGATTAGT 1560

Qy       1561 TTACGTTTCTCGTGAGAAGAGACCCGGATTTGATCACCATAAGAAGGCTGGAGCTATGAA 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 TTACGTTTCTCGTGAGAAGAGACCCGGATTTGATCACCATAAGAAGGCTGGAGCTATGAA 1620

Qy       1621 TTCCCTGATACGAGTCTCTGGGGTTCTATCAAATGCTCCTTACCTTCTGAATGTCGATTG 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TTCCCTGATACGAGTCTCTGGGGTTCTATCAAATGCTCCTTACCTTCTGAATGTCGATTG 1680
```





RESULT 2
AB214725
ID AB214725 standard; DNA; 3255 BP.
XX
AC AB214725;
XX
DT 21-JAN-2003 (first entry)
XX
DE Arabidopsis thaliana stress regulated gene SEQ ID NO 2530.
XX
KW Arabidopsis thaliana; plant; gene; stress; transgenic; ds.
XX
OS Arabidopsis thaliana.
XX
PN WO200216655-A2.
XX
PD 28-FEB-2002.
XX
PF 24-AUG-2001; 2001WO-US26685.
XX
PR 24-AUG-2000; 2000US-227866P.
PR 26-JAN-2001; 2001US-264647P.
PR 22-JUN-2001; 2001US-300111P.
XX

(SCRI) SCRIPPS RES INST.
(SYGN) SYNGENTA PARTICIPATIONS AG.

Harper JF, Kreps J, Wang X, Zhu T;

WPI; 2002-304127/34.

Identifying a stress condition to which a plant cell has been exposed and producing plants with increased tolerance to these abiotic stresses -

Claim 144; SEQ ID NO 2530; 577pp + Sequence Listing; English.

The invention relates to identifying a stress condition to which a plant cell has been exposed, comprising:
(a) contacting nucleic acid representative of expressed polynucleotides in the plant cell with an array or probes representative of the plant cell genome; and
(b) detecting a profile of expressed polynucleotides in the plant cel. characteristic of a stress response. The method is useful in the production of transgenic plants, cells and seeds and in producing plants with increased tolerance to abiotic stress. The present sequence is that of an Arabidopsis thaliana stress regulated gene (ABZ12196-ABZ17574) used in methods of the invention.
Note: The sequence data for this patent is not represented in the printed specification but is based on sequence information supplied to Derwent by the European Patent Office.

Sequence 3255 BP; 841 A; 641 C; 813 G; 960 T; 0 other;

```
Query Match            67.4%;  Score 2322;  DB 24;  Length 3255;
Best Local Similarity  99.4%;  Pred. No. 0;
Matches 3222;  Conservative    0;  Mismatches   18;  Indels    0;  Gaps    0;

Qy     8 CGGTTAATCGCCGGTTCTCACAACAGGAATGAGTTTGTCCTCATTAATGCCGATGAGAAT 67
Db    16 CGGTTAATCGCCGGTTCTCACAACAGGAATGAGTTTGTCCTCATTAATGCCGATGAGAAT 75

Qy    68 GCCCGAATAAGATCAGTCCAAGAGCTGAGTGGACAGACATGTCAAATCTGCAGAGATGAG 127
Db    76 GCCCGAATAAGATCAGTCCAAGAGCTGAGTGGACAGACATGTCAAATCTGCAGAGATGAG 135

Qy   128 ATCGAATTGACTGTTGATGGAGAACCGTTTGTGGCATGTAACGAATGTGCATTCCCTGTG 187
Db   136 ATCGAATTGACTGTTGATGGAGAACCGTTTGTGGCATGTAACGAATGTGCATTCCCTGTG 195

Qy   188 TGTAGACCTTGCTATGAGTACGAAAGACGAGAAGGCAATCAAGCTTGTCCACAGTGCAAA 247
Db   196 TGTAGACCTTGCTATGAGTACGAAAGACGAGAAGGCAATCAAGCTTGTCCACAGTGCAAA 255

Qy   248 ACCCGTTTCAAACGTCTTAAAGGAAGTCCAAGAGTTGAAGGTGATGAAGAGGAAGATGAC 307
Db   256 ACCCGTTTCAAACGTCTTAAAGGAAGTCCAAGAGTTGAAGGTGATGAAGAGGAAGATGAC 315

Qy   308 ATTGATGATTTAGACAATGAGTTTGAGTATGGAAATAATGGGATTGGATTTGATCAGGTT 367
Db   316 ATTGATGATTTAGACAATGAGTTTGAGTATGGAAATAATGGGATTGGATTTGATCAGGTT 375

Qy   368 TCTGAAGGTATGTCAATCTCTCGTCGCAACTCCGGTTTCCCACAATCTGATTTGGATTCA 427
Db   376 TCTGAAGGTATGTCAATCTCTCGTCGCAACTCCGGTTTCCCACAATCTGATTTGGATTCA 435

Qy   428 GCTCCACCTGGCTCTCAGATTCCATTGCTGACTTACGGCGACGAGGACGTTGAGATTTCT 487
Db   436 GCTCCACCTGGCTCTCAGATTCCATTGCTGACTTACGGCGACGAGGACGTTGAGATTTCT 495

Qy   488 TCTGATAGACATGCTCTTATTGTTCCTCCTTCACTTGGTGGTCATGGCAATAGAGTTCAT 547
Db   496 TCTGATAGACATGCTCTTATTGTTCCTCCTTCACTTGGTGGTCATGGCAATAGAGTTCAT 555

Qy   548 CCTGTTTCTCTTTCTGACCCGACCGTGGCTGCACATCGAAGGCTGATGGTACCTCAGAAA 607
Db   556 CCTGTTTCTCTTTCTGACCCGACCGTGGCTGCACATCCAAGGCCTATGGTACCTCAGAAA 615

Qy   608 GATCTTGCGGTTTATGGTTATGGAAGTGTCGCTTGGAAAGATCGGATGGAGGAATGGAAG 667
Db   616 GATCTTGCGGTTTATGGTTATGGAAGTGTCGCTTGGAAAGATCGGATGGAGGAATGGAAG 675

Qy   668 AGAAAGCAGAATGAGAAACTTCAGGTTGTTAGGCATGAAGGAGATCCTGATTTTGAAGAT 727
Db   676 AGAAAGCAGAATGAGAAACTTCAGGTTGTTAGGCATGAAGGAGATCCTGATTTTGAAGAT 735

Qy   728 GGTGATGATGCTGATTTTCCAATGATGGATGAGGGAAGGCAGCCATTGTCTATGAAGATA 787
Db   736 GGTGATGATGCTGATTTTCCAATGATGGATGAGGGAAGGCAGCCATTGTCTAGGAAGATA 795

Qy   788 CCAATCAAATCGAGAAAGATAAATCCTTACCGGATGTTAATTGTGCTACGTCTTGTGATT 847
Db   796 CCAATCAAATCGAGAAAGATAAATCCTTACCGGATGTTAATTGTGCTACGTCTTGTGATT 855

Qy   848 CTTGGTCTTTTCTTTCACTACCGTATTCTCCACCCGGTCAAAGATGCATATGCTTTGTGG 907
Db   856 CTTGGTCTTTTCTTTCACTACCGTATTCTCCACCCGGTCAAAGATGCATATGCTTTGTGG 915

Qy   908 CTTATTTCTGTTATATGTGAGATATGGTTTGCTGTTTCATGGGTTCTTGATCAGTTCCCT 967
Db   916 CTTATTTCTGTTATATGTGAGATATGGTTTGCTGTTTCATGGGTTCTTGATCAGTTCCCT 975

Qy   968 AAATGGTACCCTATCGAGCGAGAAACGTACTTGGACCGACTCTCATTAAGATATGAGAAA 1027
Db   976 AAATGGTACCCTATCGAGCGAGAAACGTACTTGGACCGACTCTCATTAAGATATGAGAAA 1035

Qy  1028 GAAGGGAAACCGTCGGGACTATCCCCTGTGGATGTATTTGTTAGTACAGTGGATCCATTG 1087
Db  1036 GAAGGGAAACCGTCGGGACTATCCCCTGTGGATGTATTTGTTAGTACAGTGGATCCATTG 1095

Qy  1088 AAAGAGCCTCCGCTTATTACTGCAAATACTGTCTTGTCTATTCTTGCTGTTGATTATCCT 1147
Db  1096 AAAGAGCCTCCGCTTATTACTGCAAATACTGTCTTGTCTATTCTTGCTGTTGATTATCCT 1155

Qy  1148 GTCGATAAGGTTGCTTGTTACGTATCTGATGATGGTGCTGCTATGCTTACTTTCGAAGCT 1207
Db  1156 GTCGATAAGGTTGCTTGTTACGTATCTGATGATGGTGCTGCTATGCTTACTTTCGAAGCT 1215

Qy  1208 CTTTCTGAGACCGCTGAATTCGCAAGGAAATGGGTTCCTTTCTGCAAGAAATATTGTATT 1267
Db  1216 CTTTCTGAGACCGCTGAATTCGCAAGGAAATGGGTTCCTTTCTGCAAGAAATATTGTATT 1275

Qy  1268 GAGCCTCGTGCTCCCGAATGGTATTTCTGCCATAAAATGGACTACTTGAAGAATAAAGTT 1327
Db  1276 GAGCCTCGTGCTCCCGAATGGTATTTCTGCCATAAAATGGACTACTTGAAGAATAAAGTT 1335

Qy  1328 CATCCCGCATTTGTTAGGGAGCGGCGAGCCATGAAGAGAGATTATGAAGAATTCAAAGTA 1387
Db  1336 CATCCCGCATTTGTTAGGGAGCGGCGAGCCATGAAGAGAGATTATGAAGAATTCAAAGTA 1395

Qy  1388 AAGATCAATGCTTTAGTAGCAACAGCACAGAAAGTGCCTGAGGATGGTTGGACTATGCAA 1447
Db  1396 AAGATCAATGCTTTAGTAGCAACAGCACAGAAAGTGCCTGAGGATGGTTGGACTATGCAA 1455

Qy  1448 GACGGTACACCTTGGCCCGGTAATAGTGTGCGAGATCATCCTGGCATGATTCAGGTCTTC 1507
Db  1456 GACGGTACACCTTGGCCCGGTAATAGTGTGCGAGATCATCCTGGCATGATTCAGGTCTTC 1515

Qy  1508 CTTGGAAGTGACGGTGTTCGTGATGTCGAAAACAACGAGTTGCCTCGATTAGTTTACGTT 1567
Db  1516 CTTGGAAGTGACGGTGTTCGTGATGTCGAAAACAACGAGTTGCCTCGATTAGTTTACGTT 1575

Qy  1568 TCTCGTGAGAAGAGACCCGGATTTGATCACCATAAGAAGGCTGGAGCTATGAATTCCCTG 1627
Db  1576 TCTCGTGAGAAGAGACCCGGATTTGATCACCATAAGAAGGCTGGAGCTATGAATTCCCTG 1635

Qy  1628 ATACGAGTCTCTGGGGTTCTATCAAATGCTCCTTACCTTCTGAATGTCGATTGTGATCAC 1687
Db  1636 ATACGAGTCTCTGGGGTTCTATCAAATGCTCCTTACCTTCTGAATGTCGATTGTGATCAC 1695
```

```
Qy       1688 TACATCAACAATAGCAAAGCTCTTAGAGAAGCAATGTGTTTCATGATGGATCCTCAGTCA 1747
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1696 TACATCAACAATAGCAAAGCTCTTAGAGAAGCAATGTGTTTCATGATGGATCCTCAGTCA 1755

Qy       1748 GGAAAGAAAATCTGTTATGTTCAGTTCCCTCAAAGGTTCGATGGGATTGATAGGCACGAT 1807
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1756 GGAAAGAAAATCTGTTATGTTCAGTTCCCTCAAAGGTTCGATGGGATTGATAGGCACGAT 1815

Qy       1808 CGATACTCAAATCGCAATGTTGTGTTCTTTGATATCAATATCAAAGGTTTGGATGGGCTA 1867
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1816 CGATACTCAAATCGCAATGTTGTGTTCTTTGATATCAATATCAAAGGTTTGGATGGGCTA 1875

Qy       1868 CAAGGGCCTATATACGTCGGTACAGGTTGTGTTTTCAGGAGCCAAGCGCTTTACGGATTT 1927
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1876 CAAGGGCCTATATACGTCGGTACAGGTTGTGTTTTCAGGAGCCAAGCGCTTTACGGATTT 1935

Qy       1928 GATGCACCGAAGAAGAAGAAGGGCCCACGTAAGACATGCAATTGCTGGCCAAAATGGTGT 1987
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1936 GATGCACCGAAGAAGAAGAAGGGCCCACGTAAGACATGCAATTGCTGGCCAAAATGGTGT 1995

Qy       1988 CTCCTATGTTTTGGTTCAAGAAAGAATCGTAAAGCAAAGACAGTGGCTGCGGATAAGAAG 2047
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1996 CTCCTATGTTTTGGTTCAAGAAAGAATCGTAAAGCAAAGACAGTGGCTGCGGATAAGAAG 2055

Qy       2048 AAGAAGAATAGGGAAGCGTCAAAGCAGATCCACGCATTAGAAAATATCGAAGAGGGCCGC 2107
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2056 AAGAAGAATAGGGAAGCGTCAAAGCAGATCCACGCATTAGAAAATATCGAAGAGGGCCGC 2115

Qy       2108 GGTCATAAAGTTCTTAACGTAGAACAGTCAACCGAGGCAATGCAAATGAAGTTGCAGAAG 2167
             |    ||||| |  ||||||||| |||||||||||||||||||||||||||||| |||||
Db       2116 GTCACTAAAGGTTCTAACGTAGAACAGTCAACCGAGGCAATGCAAATGAAGTTGGAGAAG 2175

Qy       2168 AAATATGGGCAGTCTCCTGTATTTGTTGCATCTGCGCGTCTGGAGAATGGTGGGATGGCT 2227
             |||| |||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db       2176 AAATTTGGGCAGTCTCCTGTATTTGTTGCATCTGCGCGTATGGAGAATGGTGGGATGGCT 2235

Qy       2228 AGAAACGCAAGCCCGGCTTGTCTGCTTAAAGAAGCCATCCAAGTCATTAGTCGCGGATAT 2287
             ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||| ||||
Db       2236 AGAAACGCAAGCCCGGCTTGTCTGCTTAAAGAAGCCATCCAAGTCATTAGTTGCGGATAT 2295

Qy       2288 GAAGATAAAACTGAATGGGGAAAAGAGATTGGGTGGATCTATGGTTCTGTTACCGAAGAT 2347
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2296 GAAGATAAAACTGAATGGGGAAAAGAGATTGGGTGGATCTATGGTTCTGTTACCGAAGAT 2355

Qy       2348 ATTCTTACGGGTTCTAAGATGCATTCTCATGGTTGGAGACATGTTTATTGTACACCAAAG 2407
             |||||||||||| |||||||||||||||||||||||| |||||||||||||||||||||||
Db       2356 ATTCTTACGGGTTTTAAGATGCATTCTCATGGTTGGAGATCTGTTTATTGTACACCAAAG 2415

Qy       2408 TTAGCGGCTTTCAAAGGATCAATTGCAATCAATGTTTCGGAAGTCTCCATCAAGTTCTT 2467
             |||| |||||||||||||||  ||| |||| ||| |||| |||||||||||||||||||||
Db       2416 TTAGCGGCTTTCAAAGGATCAATTGCAATCAATGTTTCGGAAGTCTCCATCAAGTTCTT 2475

Qy       2468 CGATGGGCGCTTGGGTCGGTTGAGATTTTCTTGAGTAGGCATGTCCTATTTGGTATGGT 2527
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2476 CGATGGGCGCTTGGGTCGGTTGAGATTTTCTTGAGTAGGCATGTCCTATTTGGTATGGT 2535

Qy       2528 TATGGAGGTGGGTTGAAATGGCTTGAGCGGTTGTCCTACATTAACTCTGTGGTTTACCCG 2587
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2536 TATGGAGGTGGGTTGAAATGGCTTGAGCGGTTGTCCTACATTAACTCTGTGGTTTACCCG 2595

Qy       2588 TGGACCTCTCTACCGCTCATCGTTTACTGTTCTCTCCCTGCTATCTGTCTTCTCACTGGA 2647
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2596 TGGACCTCTCTACCGCTCATCGTTTACTGTTCTCTCCCTGCTATCTGTCTTCTCACTGGA 2655

Qy       2648 AAATTCATCGTTCCCGAGATTAGCAACTATGCGAGTATCCTCTTCATGGCGCTCTTCTCG 2707
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2656 AAATTCATCGTTCCCGAGATTAGCAACTATGCGAGTATCCTCTTCATGGCGCTCTTCTCG 2715

Qy       2708 TCGATTGCAATAACGGGTATTCTCGAGATGCAATGGGGCAAAGTTGGGATCGATGATTGG 2767
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2716 TCGATTGCAATAACGGGTATTCTCGAGATGCAATGGGGCAAAGTTGGGATCGATGATTGG 2775

Qy       2768 TGGAGAAACGAACAGTTTTGGGTCATTGGAGGTGTTTCTGCGCATCTGTTTGCTCTCTTC 2827
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2776 TGGAGAAACGAACAGTTTTGGGTCATTGGAGGTGTTTCTGCGCATCTGTTTGCTCTCTTC 2835

Qy       2828 CAAGGTCTCCTCAAGGTTCTTGCTGGTGTCGACACTAACTTCACAGTCACATCAAAAGCA 2887
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2836 CAAGGTCTCCTCAAGGTTCTTGCTGGTGTCGACACTAACTTCACAGTCACATCAAAAGCA 2895

Qy       2888 GCTGATGATGGAGAGTTCTCTGACCTTTACCTCTTCAAATGGACTTCACTTCTCATCCCT 2947
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2896 GCTGATGATGGAGAGTTCTCTGACCTTTACCTCTTCAAATGGACTTCACTTCTCATCCCT 2955

Qy       2948 CCAATGACTCTACTCATCATAAACGTCATTGGAGTCATAGTCGGAGTCTCTGATGCCATC 3007
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2956 CCAATGACTCTACTCATCATAAACGTCATTGGAGTCATAGTCGGAGTCTCTGATGCCATC 3015

Qy       3008 AGCAATGGATACGACTCGTGGGGACCGCTTTTCGGAAGACTGTTCTTTGCACTTTGGGTC 3067
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3016 AGCAATGGATACGACTCGTGGGGACCGCTTTTCGGAAGACTGTTCTTTGCACTTTGGGTC 3075

Qy       3068 ATCATTCATCTTTACCCGTTCCTTAAAGGTTTGCTTGGGAAACAAGATAGAATGCCAACC 3127
             |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db       3076 ATCATTCATCTTTACCCGTTCCTTAAAGGTTTGCTTGGGAAACAAGATAGAATGCCAACC 3135

Qy       3128 ATTATTGTCGTCTGGTCCATCCTCCTGGCCTCGATTCTTACACTTCTTTGGGTCCGGGTT 3187
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3136 ATTATTGTCGTCTGGTCCATCCTCCTGGCCTCGATTCTTACACTTCTTTGGGTCCGGGTT 3195

Qy       3188 AATCCGTTTGTGGCGAAAGGCGGTCCTATTCTCGAGATCTGTGGTTTAGACTGCTTGTGA 3247
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3196 AATCCGTTTGTGGCGAAAGGCGGTCCTATTCTCGAGATCTGTGGTTTAGACTGCTTGTGA 3255
```

RESULT 3
ABL93998/c
ID ABL93998 standard; cDNA; 447 BP.
XX
AC ABL93998;
XX
DT 10-JUN-2002 (first entry)
XX
DE Arabidopsis thaliana nucleic acid sequence Ref:2027763 SEQ ID NO:763.
XX
KW Arabidopsis thaliana; insecticide; fungicide; plant; mapping; diagnosis;
KW genetic modification; gene; ss.
XX
OS Arabidopsis thaliana.
XX
PN US2002023280-A1.
XX
PD 21-FEB-2002.
XX
PF 26-JAN-2001; 2001US-0770444.
XX
PR 27-JAN-2000; 2000US-178502P.
XX
PA (GORL/) GORLACH J.
PA (ANYY/) AN Y.
PA (HAMI/) HAMILTON C M.
PA (PRIC/) PRICE J L.
PA (RAIN/) RAINES T M.
PA (YUYY/) YU Y.
PA (RAME/) RAMEAKA J G.
PA (PAGE/) PAGE A.
PA (MATH/) MATHEW A V.
PA (LEDF/) LEDFORD B L.
PA (WOES/) WOESSNER J P.
PA (HAAS/) HAAS W D.
PA (GARC/) GARCIA C A.
PA (KRIC/) KRICKER M.
PA (SLAT/) SLATER T.
PA (DAVI/) DAVIS K R.
PA (ALLE/) ALLEN K.
PA (HOFF/) HOFFMAN N.

A (HURB/) HURBAN P.
X
I Gorlach J, An Y, Hamilton CM, Price JL, Raines TM, Yu Y;
I Rameaka JG, Page A, Mathew AV, Ledford BL, Woessner JP, Haas WD;
I Garcia CA, Kricker M, Slater T, Davis KR, Allen K, Hoffman N;
I Hurban P;
K
R WPI; 2002-267486/31.
X
T New Arabidopsis thaliana nucleic acid, for identifying homologous
T genes, producing compositions that modulate the expression or function
T of its encoded protein, and mapping functional regions of a protein
X
S Claim 1; SEQ ID 763; 44pp; English.
X
C The present invention describes an Arabidopsis thaliana nucleic acid (I)
C comprising a sequence capable of hybridising under stringent conditions
C to a sequence (S1) selected from any one of the 999 sequences given in
C ABL93236 to ABL94234. (I) have insecticide and fungicide activities, and
C they can be used as protein expression modulators. (I) can be used in
C identifying homologous or related genes, in producing compositions that
C modulate the expression or function of their encoded proteins, mapping
C functional regions of the proteins, and in studying associated
C physiological pathways. (I) can also be used: (1) for the genetic
C manipulation of cells, particularly plant cells; (2) in screening assays
C of various plant strains to determine the strains that are best capable
C of withstanding a particular disease or environmental stress; (3) for
C enhancing or inhibiting production of a biosynthetic product in a plant;
C (4) as probes in mapping and in diagnosis, in genetic modification and
C for screening purposes, to generate additional copies of the nucleic
C acids, to generate ribozymes or antisense oligonucleotides, and as
C single-stranded DNA probes or as triple-stand forming oligonucleotides;
C and (5) for generating genetically modified transgenic organisms.
C Note: The sequence data for this patent did not form part of the printed
C specification, but was obtained in electronic format directly from the
C USPTO web site.
X
Q Sequence 447 BP; 156 A; 105 C; 62 G; 124 T; 0 other;

```
 Query Match             10.5%;  Score 361;  DB 24;  Length 447;
 Best Local Similarity   100.0%;  Pred. No. 4e-170;
 Matches  361;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3076 TCTTTACCCGTTCCTTAAAGGTTTGCTTGGGAAACAAGATAGAATGCCAACCATTATTGT 3135
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        447 TCTTTACCCGTTCCTTAAAGGTTTGCTTGGGAAACAAGATAGAATGCCAACCATTATTGT 388

Qy       3136 CGTCTGGTCCATCCTCCTGGCCTCGATTCTTACACTTCTTTGGGTCCGGGTTAATCCGTT 3195
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        387 CGTCTGGTCCATCCTCCTGGCCTCGATTCTTACACTTCTTTGGGTCCGGGTTAATCCGTT 328

Qy       3196 TGTGGCGAAAGGCGGTCCTATTCTCGAGATCTGTGGTTTAGACTGCTTGTGATTCGATTG 3255
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        327 TGTGGCGAAAGGCGGTCCTATTCTCGAGATCTGTGGTTTAGACTGCTTGTGATTCGATTG 268

Qy       3256 ACCGGTGGATGGGTTGGTGAAAAAGGTTTAATTCCCACGGATCAAAGAGAGGTAAGAGAG 3315
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        267 ACCGGTGGATGGGTTGGTGAAAAAGGTTTAATTCCCACGGATCAAAGAGAGGTAAGAGAG 208

Qy       3316 ATATTGTTTTACCTCTAAAAGACTCCTTCATTGTGTTCATTAGATGATGAAAAATGAAAA 3375
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        207 ATATTGTTTTACCTCTAAAAGACTCCTTCATTGTGTTCATTAGATGATGAAAAATGAAAA 148

Qy       3376 GAAAAAGAAGATTTAATTTTGTTACGAGAATTGTTATTTTTGCAAGAATGTGTTGTAGAT 3435
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        147 GAAAAAGAAGATTTAATTTTGTTACGAGAATTGTTATTTTTGCAAGAATGTGTTGTAGAT 88

Qy       3436 A 3436
              |
Db         87 A 87
```

RESULT 4

ABX61330/c
ID ABX61330 standard; DNA; 342 BP.
XX
AC ABX61330;
XX
DT 26-FEB-2003 (first entry)
XX
DE Arabidopsis thaliana polynucleotide #676
XX
KW Thale cress; gene; ds; genetic manipulation; plant; biosynthesis;
KW genetic modification; environmental stress; disease resistance;
KW fungicide; insecticide; stress tolerance
XX
OS Arabidopsis thaliana
XX
PN US2002142319-A1
XX
PD 03-OCT-2002.
XX
PF 07-AUG-2001; 2001US-0923655.
XX
PR 13-AUG-1999; 99US-148784P.
PR 11-AUG-2000; 2000US-0638258.
XX
PA (GORL/) GORLACH J.
PA (ANYY/) AN Y.
PA (HAMI/) HAMILTON C M.
PA (PRIC/) PRICE J L.
PA (HARG/) HARGISS T R.
PA (YUYY/) YU Y.
PA (RAME/) RAMEAKA J G.
PA (PAGE/) PAGE A.
PA (MATH/) MATHEW A V.
PA (LEDF/) LEDFORD B L.
PA (WOES/) WOESSNER J P.
PA (HAAS/) HAAS W D.
PA (GARC/) GARCIA C A.
XX
PI Gorlach J, An Y, Hamilton CM, Price JL, Hargiss TR, Yu Y;
PI Rameaka JG, Page A, Mathew AV, Ledford BL, Woessner JP, Haas WD;
PI Garcia CA;
XX
DR WPI; 2003-102509/09.
XX
PT Novel Arabidopsis thaliana nucleic acid useful for constructing a
PT transgenic plant with enhanced disease resistance and enhanced traits
PT of interest, as probes, and in diagnosis and screening purposes
XX
PS Claim 1; Page 208; 277pp; English.
XX
CC The invention relates to Arabidopsis thaliana nucleic acid sequences. The
CC DNA sequences and the polypeptides they encode are useful for identifying
CC homologous or related genes, for producing compositions that modulate the
CC expression or function of the polypeptides, for mapping functional
CC regions of the protein, in diagnosis, for studying associated
CC physiological pathways, for genetic manipulation of cells, preferably
CC plant cells, in screening assays of various plant strains to determine
CC the strains that are capable of withstanding a particular disease or
CC environmental stress, for enhancing or inhibiting production of
CC biosynthetic products in plants and to create genetically modified and
CC transgenic organisms, such as plant cells and plants. Transgenic plants
CC are useful for introducing or improving disease resistance and stress
CC tolerance in plants, screening biologically active agents, such as
CC fungicides and insecticides, and for identifying factors involved in
CC biosynthetic pathways of nutritional, commercial or medicinal value.
CC Sequences ABX60655-ABX61554 represent Arabidopsis thaliana
CC polynucleotides of the invention.
XX
SQ Sequence 342 BP; 107 A; 86 C; 50 G; 96 T; 3 other;

```
 Query Match             6.4%;  Score 222;  DB 25;  Length 342;
 Best Local Similarity   100.0%;  Pred. No. 1.9e-100;
 Matches  222;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy       3215 ATTCTCGAGATCTGTGGTTTAGACTGCTTGTGATTCGATTGACCGGTGGATGGGTTGGTG 3274
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        265 ATTCTCGAGATCTGTGGTTTAGACTGCTTGTGATTCGATTGACCGGTGGATGGGTTGGTG 206

Qy       3275 AAAAAGGTTTAATTCCCACGGATCAAAGAGAGGTAAGAGAGATATTGTTTTACCTCTAAA 3334
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        205 AAAAAGGTTTAATTCCCACGGATCAAAGAGAGGTAAGAGAGATATTGTTTTACCTCTAAA 146

Qy       3335 AGACTCCTTCATTGTGTTCATTAGATGATGAAAAATGAAAAGAAAAAGAAGATTTAATTT 3394
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        145 AGACTCCTTCATTGTGTTCATTAGATGATGAAAAATGAAAAGAAAAAGAAGATTTAATTT 86

Qy       3395 TGTTACGAGAATTGTTATTTTTGCAAGAATGTGTTGTAGATA 3436
             |||||||||||||||||||||||||||||||||||||||||||
Db         85 TGTTACGAGAATTGTTATTTTTGCAAGAATGTGTTGTAGATA 44


RESULT 5
AAZ58270
ID   AAZ58270 standard; cDNA; 2890 BP.
XX
AC   AAZ58270;
XX
DT   08-MAY-2000  (first entry)
XX
DE   Soybean cellulose synthase cDNA clone ssl.pk0036.c10.
XX
KW   Soybean; cellulose synthase; transgenic plant; ss.
XX
OS   Glycine max.
XX
FH   Key             Location/Qualifiers
FT   CDS             2..2383
FT                   /*tag= a
FT                   /partial
XX
PN   WO200034166-A2.
XX
PD   27-JAN-2000.
XX
PF   13-JUL-1999;   99WO-US15871.
XX
PR   14-JUL-1998;   98US-0092844.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I
XX
PI   Allen SM,  Fader GM,  Falco SC,  Kinney AJ,  Lightner JE,  Miao G;
PI   Rafalski JA,  Thorpe CJ;
XX
DR   WPI; 2000-182431/16.
DR   P-PSDB; AAY58839.
XX
PT   Nucleic acid fragments encoding cellulose biosynthetic enzyme useful as
PT   probes for isolating cDNAs and genes encoding homologous proteins, for
PT   producing transgenic plants -
XX
PS   Claim 3; Page 62-63; 93pp; English.
XX
CC   The present sequence is that of cDNA clone ssl.pk0036.c10 encoding
CC   a portion (see AAY58839) of soybean cellulose synthase (CS). The
CC   cDNA clone was isolated from a cDNA library constructed from
CC   soybean seedling (5-10 day) tissue on the basis of homology to
CC   Arabidopsis and cotton CS sequences. The invention relates to
CC   isolated nucleic acid fragments encoding plant CS and to CS
CC   polypeptides. It also relates to the construction of a chimeric
CC   gene encoding all or a portion of the CS, in sense or antisense
CC   orientation, where expression of the gene results in altered levels
CC   of the CS in transformed host cells. The host cells can be used to
CC   screen compounds for their ability to inhibit CS activity. CS
CC   nucleic acids are also useful for producing transgenic plants having
CC   altered levels of CS, and hence altered levels of fibre. CS may
CC   also serve as a target for the development of novel herbicides.
XX
SQ   Sequence 2890 BP; 808 A; 516 C; 695 G; 871 T; 0 other;

  Query Match             0.9%;  Score 32;  DB 21;  Length 2890;
  Best Local Similarity   100.0%;  Pred. No. 3.2e-05;
  Matches   32;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1481 GATCATCCTGGCATGATTCAGGTCTTCCTTGG 1512
             ||||||||||||||||||||||||||||||||
Db        614 GATCATCCTGGCATGATTCAGGTCTTCCTTGG 645


RESULT 6
ABZ12754
ID   ABZ12754 standard; DNA; 3255 BP.
XX
AC   ABZ12754;
XX
DT   21-JAN-2003  (first entry)
XX
DE   Arabidopsis thaliana stress regulated gene SEQ ID NO 559.
XX
KW   Arabidopsis thaliana; plant; gene; stress; transgenic; ds.
XX
OS   Arabidopsis thaliana.
XX
PN   WO200216655-A2.
XX
PD   28-FEB-2002.
XX
PF   24-AUG-2001; 2001WO-US26685.
XX
PR   24-AUG-2000; 2000US-227866P.
PR   26-JAN-2001; 2001US-264647P.
PR   22-JUN-2001; 2001US-300111P.
XX
PA   (SCRI ) SCRIPPS RES INST.
PA   (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX
PI   Harper JF,  Kreps J,  Wang X,  Zhu T;
XX
DR   WPI; 2002-304127/34.
XX
PT   Identifying a stress condition to which a plant cell has been exposed
PT   and producing plants with increased tolerance to these abiotic stresses
PT   -
XX
PS   Claim 144; SEQ ID NO 559; 577pp + Sequence Listing; English.
XX
CC   The invention relates to identifying a stress condition to which a plant
CC   cell has been exposed, comprising:
CC   (a) contacting nucleic acid representative of expressed polynucleotides
CC   in the plant cell with an array or probes representative of the plant
CC   cell genome; and
CC   (b) detecting a profile of expressed polynucleotides in the plant cell
CC   characteristic of a stress response. The method is useful in the
CC   production of transgenic plants, cells and seeds and in producing plants
CC   with increased tolerance to abiotic stress. The present sequence is that
CC   of an Arabidopsis thaliana stress regulated gene (ABZ12196-ABZ17574) used
CC   in methods of the invention.
CC   Note: The sequence data for this patent is not represented in the printed
CC   specification but is based on sequence information supplied to Derwent by
CC   the European Patent Office.
XX
SQ   Sequence 3255 BP; 884 A; 621 C; 798 G; 952 T; 0 other;

  Query Match             0.9%;  Score 32;  DB 24;  Length 3255;
  Best Local Similarity   100.0%;  Pred. No. 3.2e-05;
  Matches   32;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        419 TTGGATTCAGCTCCACCTGGCTCTCAGATTCC 450
              ||||||||||||||||||||||||||||||||
Db        415 TTGGATTCAGCTCCACCTGGCTCTCAGATTCC 446
```

```
ESULT 7
AV06566
D   AAV06566 standard; cDNA; 3828 BP.
X
C   AAV06566;
X
T   06-JUL-1998  (first entry)
X
E   Arabidopsis cellulose biosynthetic gene clone Ath-A.
X
W   Cellulose; cellulose synthase; RSW1 gene; beta-1,4-glucan;
W   transgenic plant; ss.
X
S   Arabidopsis thaliana var. Columbia.
X
H   Key             Location/Qualifiers
T   CDS             239..3490
T                   /*tag= a
X
N   WO9800549-A1.
X
D   08-JAN-1998.
X
F   24-JUN-1997;   97WO-AU00402.
X
R   27-JUN-1996;   96AU-0000699.
X
A   (CSIR ) COMMONWEALTH SCI & IND RES ORG.
A   (AUSU ) UNIV AUSTRALIAN NAT.
X
I   Arioli A,  Betzner AS,  Peng L,  Williamson RE;
X
R   WPI; 1998-086974/08.
R   P-PSDB; AAW33818.
X
T   DNA encoding cellulose biosynthetic enzyme - useful for manipulation
T   of cellulose and beta-1,4-glucan
X
S   Claim 12; Page 123-130; 207pp; English.
X
C   cDNA clone Ath-A was isolated from an Arabidopsis thaliana cDNA
C   library using probes produced by PCR with primers (see AAT99632-34)
C   based on cellulose synthase RSW1 genomic clone 23H12 (see AAV06563)
C   and EST clone AAT20782 (see AAV06562).  It is closely related to RSW1
C   cDNA (see AAV06565) and to partial genomic clone 12C4 (see AAV06564),
C   and encodes a 1084-polypeptide (see AAW33818).  Claimed nucleic acid
C   molecules (see AAV06562-69) coding for claimed polypeptides (see
C   AAW33816-20 and AAW46202) involved in cellulose biosynthesis can be
C   used to manipulate the cellulose and/or beta-glucan content of
C   transgenic plants.  Expression of nucleic acids in the sense
C   orientation increases the level of cellulose and reduces the level
C   of non-crystalline beta-1,4-glucan and starch, providing plants
C   with modified strength and/or shape and/or fibre properties, or
C   having increased resistance to stresses or pests.  Antisense,
'C   ribozyme or co-suppression molecules can be used to reduce the
:C   cellulose content of a transgenic plant, e.g. to improve
:C   digestibility or to alter carbon partitioning such that increased
:C   carbon is available for growth, rather than deposited as cellulose.
XX
3Q  Sequence 3828 BP; 1056 A; 721 C; 912 G; 1139 T; 0 other;

 Query Match             0.9%;  Score 32;  DB 19;  Length 3828;
 Best Local Similarity   100.0%;  Pred. No. 3.2e-05;
 Matches   32;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        419 TTGGATTCAGCTCCACCTGGCTCTCAGATTCC 450
              ||||||||||||||||||||||||||||||||
Db        653 TTGGATTCAGCTCCACCTGGCTCTCAGATTCC 684


RESULT 8
```

```
AAV06564
ID   AAV06564 standard; DNA; 5009 BP.
XX
AC   AAV06564;
XX
DT   06-JUL-1998  (first entry)
XX
DE   Arabidopsis cellulose synthase RSW1 gene variant clone 12C4.
XX
KW   Beta-glycosyltransferase; cellulose synthase; RSW1 gene;
KW   beta-1,4-glucan; transgenic plant; ss.
XX
OS   Arabidopsis thaliana var. Columbia
XX
FH   Key             Location/Qualifiers
FT   promoter        1..296
FT                   /*tag= c
FT   TATA signal     [illegible]
FT                   /*tag= d
FT   CDS             [illegible]
FT                   /*tag= a
FT                   /note= "contains introns"
FT   exon            863..964
FT                   /*tag= d
FT                   /number= 1
FT   intron          964..1464
FT                   /*tag= e
FT                   /number= 1
FT   exon            1464..1640
FT                   /*tag= f
FT                   /number= 2
FT   intron          1641..1922
FT                   /*tag= g
FT                   /number= 2
FT   exon            1923..2025
FT                   /*tag= h
FT                   /number= 3
FT   intron          2026..2121
FT                   /*tag= i
FT                   /number= 3
FT   exon            2122..2311
FT                   /*tag= j
FT                   /number= 4
FT   intron          2312..2420
FT                   /*tag= k
FT                   /number= 4
FT   exon            2421..2687
FT                   /*tag= l
FT                   /number= 5
FT   intron          2688..2775
FT                   /*tag= m
FT                   /number= 5
FT   exon            2776..3121
FT                   /*tag= n
FT                   /number= 6
FT   intron          3122..3219
FT                   /*tag= o
FT                   /number= 6
FT   exon            3220..3357
FT                   /*tag= p
FT                   /number= 7
FT   intron          3358..3506
FT                   /*tag= q
FT                   /number= 7
FT   exon            3507..3623
FT                   /*tag= r
FT                   /number= 8
FT   intron          3624..3722
FT                   /*tag= s
FT                   /number= 8
FT   exon            3723..3935
FT                   /*tag= t
FT                   /number= 9
```

```
FT   intron          3936..4026
FT                   /*tag= u
FT                   /number= 9
FT   exon            4027..4297
FT                   /*tag= v
FT                   /number= 10
FT   intron          4298..4379
FT                   /*tag= w
FT                   /number= 10
FT   exon            4380..4576
FT                   /*tag= x
FT                   /number= 11
FT   intron          4577..4658
FT                   /*tag= y
FT                   /number= 11
FT   exon            4659..5009
FT                   /*tag= z
FT                   /number= 12
FT                   /note= "5' portion of exon 12"
XX
PN   WO9800549-A1.
XX
PD   08-JAN-1998
XX
PF   24-JUN-1997;    97WO-AU00402.
XX
PR   27-JUN-1996;    96AU-0000699.
XX
PA   (CSIR ) COMMONWEALTH SCI & IND RES ORG.
PA   (AUSU ) UNIV AUSTRALIAN NAT.
XX
PI   Arioli A,  Betzner AS,  Peng L,  Williamson RE;
XX
DR   WPI; 1998-086974/08.
DR   P-PSDB; AAW33818.
XX
PT   DNA encoding cellulose biosynthetic enzyme - useful for manipulation
PT   of cellulose and beta-1,4-glucan
XX
PS   Claim 12; Page 104-109; 207pp; English.
XX
CC   This DNA molecule comprises cosmid clone 12C4, a partial clone
CC   comprising a genomic gene variant of RSW1 (see AAV06563) that encodes
CC   the cellulose synthase of Arabidopsis thaliana. Clone 12C4 was
CC   isolated by hybridisation using EST AAT20782 (see AAV06562) insert DNA.
CC   It is equivalent (but 3' truncated) to cDNA clone Ath-A (see
CC   AAV06566). Claimed nucleic acid molecules (see AAV06562-69) coding for
CC   claimed polypeptides (see AAW33816-20 and AAW46202) involved in
CC   cellulose biosynthesis can be used to manipulate the cellulose
CC   and/or beta-glucan content of transgenic plants. Expression of
CC   nucleic acids in the sense orientation increases the level of
CC   cellulose and reduces the level of non crystalline beta 1,4-glucan
CC   and starch, providing plants with modified strength and/or shape
CC   and/or fibre properties, or having increased resistance to stresses
CC   or pests. Antisense, ribozyme or co-suppression molecules can be
CC   used to reduce the cellulose content of a transgenic plant, e.g. to
CC   improve digestibility or to alter carbon partitioning such that
CC   increased carbon is available for growth, rather than deposited as
CC   cellulose.
XX
SQ   Sequence 5009 BP; 1352 A; 901 C; 1048 G; 1696 T; 12 other;

  Query Match             0.9%;  Score 32;  DB 19;  Length 5009;
  Best Local Similarity   100.0%;  Pred. No. 3.2e-05;
  Matches   32;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        419 TTGGATTCAGCTCCACCTGGCTCTCAGATTCC 450
              ||||||||||||||||||||||||||||||||
Db       1787 TTGGATTCAGCTCCACCTGGCTCTCAGATTCC 1818


RESULT 9
AAV08372
ID   AAV08372 standard; cDNA to mRNA; 3207 BP.
XX
AC   AAV08372;
XX
DT   08-FEB-1999  (first entry)
XX
DE   Cellulose synthase PcsA1 coding sequence.
XX
KW   Cellulose synthase; cotton; ds.
XX
OS   Gossypium hirsutum.
XX
FH   Key             Location/Qualifiers
FT   CDS             77..3001
FT                   /*tag= a
XX
PN   EP875575-A2.
XX
PD   04-NOV-1998.
XX
PF   31-MAR-1998;    98EP-0302489.
XX
PR   01-APR-1997;    97JP-0083133.
XX
PA   (NISN ) NISSHINBO IND INC.
XX
PI   Aotsuka S,  Hasegawa O,  Hayashi T,  Ihara Y;
XX
DR   WPI; 1998-559440/48.
DR   P-PSDB; AAW73308.
XX
PT   New DNA encoding cotton cellulose synthase proteins - useful for
PT   controlling cellulose synthesis in prokaryotic or eukaryotic cells
XX
PS   Example 5; Page 14-18; 51pp; English.
XX
CC   This sequence encodes a cellulose synthase of the invention. The DNA can
CC   be used for controlling cellulose synthesis in a cell by introducing the
CC   DNA into the cell, expressing RNA with a nucleotide sequence homologous
CC   to the DNA or a nucleotide sequence complementary to the DNA. The cells
CC   are prokaryotic cells i.e. acetobacterium and/or eukaryotic cells such as
CC   yeasts i.e. Saccharomyces spp., plant cells (cotton plants) and cultured
CC   cells of mammals etc.
XX
SQ   Sequence 3207 BP; 883 A; 662 C; 723 G; 939 T; 0 other;

  Query Match             0.8%;  Score 29;  DB 19;  Length 3207;
  Best Local Similarity   100.0%;  Pred. No. 0.001;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1070 AGTACAGTGGATCCATTGAAAGAGCCTCC 1098
              |||||||||||||||||||||||||||||
Db        845 AGTACAGTGGATCCATTGAAAGAGCCTCC 873


RESULT 10
AAV34432
ID   AAV34432 standard; cDNA; 3328 BP.
XX
AC   AAV34432;
XX
DT   28-SEP-1998  (first entry)
XX
DE   Cotton cellulose synthase CelA1 cDNA.
XX
KW   Cellulose synthase; CelA1 gene; cotton; forest tree; plant fibre;
KW   wood; transgenic plant; ds
XX
OS   Gossypium hirsutum Acala SJ-2.
XX
FH   Key             Location/Qualifiers
FT   CDS             179..3103
FT                   /*tag= a
```

X
N WO9818949-A2.
X
D 07-MAY-1998.
X
F 29-OCT-1997; 97WO-US19529.
X
R 29-OCT-1996; 96US-0029987.
X
A (CALJ) CALGENE INC.
X
I Delmer DP, Pear JR, Stalker DM;
X
R WPI; 1998-272238/24.
R P-PSDB; AAW60039.
X
T New isolated plant cellulose synthesis enzyme DNA - is used to
T develop products for producing modified plant fibre products,
T particularly for modifying cotton fibre and wood quality
X
S Claim 5; Fig 6A-E; 49pp; English.
X
C This full-length cDNA sequence corresponds to the cotton CelA1 gene
C transcript and encodes a 973-amino acid cellulose synthase (see
C AAW60039). 2 cDNA clones that exhibited small blocks of amino acid
C homology to proteins encoded by bacterial CelA genes were isolated
C from an unamplified cDNA library prepared from mRNA of fibres of
C cotton Acala SJ-2 harvested at 21 dpa (time of maximal secondary
C wall cellulose synthesis). These 2 clones were partially homologous
C at the nucleotide and amino acid levels and were designated CelA1
C and CelA2 (see AAV34433), respectively. The 2 cotton CelA genes are
C both expressed in developing fibre at approximately 500 times their
C level of expression in other cotton tissues, and constitute
C approximately 1-2% of the 24 dpa fibre mRNA. The CelA1 gene
C promoter region (see AAV34434) is also claimed. The isolated cDNA
C clones are used in claimed methods for modifying fibre phenotype
C in a cotton plant and for modifying wood quality phenotype in a
C forest tree species. Lignin can be reduced by over-expressing the
C cellulose synthase product and diverting production into cellulose.
X
Q Sequence 3328 BP; 914 A; 690 C; 763 G; 961 T; 0 other;

Query Match 0.8%; Score 29; DB 19; Length 3328;
Best Local Similarity 100.0%; Pred. No. 0.001;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

y 1070 AGTACAGTGGATCCATTGAAAGAGCCTCC 1098
||||||||||||||||||||||||||||
b 947 AGTACAGTGGATCCATTGAAAGAGCCTCC 975

ESULT 11
AZ99533
D AAZ99533 standard; DNA; 3704 BP.
X
C AAZ99533;
X
T 03-JUL-2000 (first entry)
X
E DNA encoding a maize cellulose synthase.
X
KW Maize; cellulose synthase; stalk quality; stand; silage; cellulose;
KW transgenic plant; plant breeding marker; ss.
X
OS Zea mays.
X
FH Key Location/Qualifiers
FT CDS 272..3496
FT /*tag= a
FT /product= "cellulose synthase"
FT /transl_except= (pos: 1544..1546, aa: Xaa)
FT /note= "no termination codon given; Xaa is an
FT unspecified amino acid"

XX
PN WO200009706-A2.
XX
PD 24-FEB 2000.
XX
PF 16-AUG-1999; 99WO US18760.
XX
PR 17-AUG 1998; 98US-0096822
XX
PA (PION-) PIONEER HI BRED INT INC
XX
PI Dhugga KS, Helentjaris TG, Bowen BA, Wang X;
XX
DR WPI; 2000-224343/19.
DR P-PSDB; AAY84121
XX
PT New genes which encode maize cellulose synthase polypeptides in plants
PT useful for modulating the expression of cellulose synthase in plants
PT and to produce transgenic plants expressing the novel protein
XX
PS Claim 1; Page 191-196; 219pp; English
XX
CC The present sequence encodes a maize cellulose synthase polypeptide.
CC The cellulose synthase can be used for the improvement of stalk quality
CC for improved stand or silage. It also provides an increased concentration
CC of cellulose in the pericarp, hardening the kernel and improving its
CC handling ability. The sequences are used to produce transgenic plants
CC and seeds expressing the cellulose synthase. The polynucleotide is
CC used for modulating, preferably increasing, the level of the synthase
CC in a plant cell. The plants are preferably monocots. The polynucleotide
CC is also used as a probe or primer in the detection quantitation or
CC isolation of gene transcripts. The probes are useful in detecting
CC deficiencies in the level of mRNA in screenings for desired transgenic
CC plant, for detecting mutations in the gene, for monitoring upregulation
CC of expression or changes in enzyme activity in screening assays of
CC compounds, for detection of any number of allelic variants of the gene,
CC or for use as molecular markers in plant breeding programs. The
CC isolated nucleic acids of the present invention can also be used for
CC recombinant expression of their encoded polypeptides or for use as
CC immunogens in the preparation and/or screening of antibodies. The
CC proteins can be employed in assays for enzyme agonists or antagonists
CC of enzyme function or for use of immunogens or antigens to obtain
CC antibodies specifically immunoreactive with a protein.
XX
SQ Sequence 3704 BP; 909 A; 838 C; 962 G; 994 T; 1 other;

Query Match 0.8%; Score 29; DB 21; Length 3704;
Best Local Similarity 100.0%; Pred. No. 0.001;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2312 GAGATTGGGTGGATCTATGGTTCTGTTAC 2340
||||||||||||||||||||||||||||
Db 2564 GAGATTGGGTGGATCTATGGTTCTGTTAC 2592

RESULT 12
AAZ58268
ID AAZ58268 standard; cDNA; 3517 BP.
XX
AC AAZ58268;
XX
DT 08-MAY 2000 (first entry)
XX
DE Soybean cellulose synthase cDNA clone sdp2c.pk005.o22.
XX
KW Soybean; cellulose synthase; transgenic plant; ss.
XX
OS Glycine max.
XX
FH Key Location/Qualifiers
FT CDS 158..3278
FT /*tag= a
FT /transl_except= (pos:758..761, aa:Xaa)

```
FT                   /note= "apparent 1 base addition, alters the
FT                         reading frame"
XX
PN   WO200004166-A2.
XX
PD   27-JAN-2000.
XX
PF   13-JUL-1999;   99WO-US15871.
XX
PR   14-JUL-1998;   98US-0092844.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I
XX
PI   Allen SM,  Fader GM,  Falco SC,  Kinney AJ,  Lightner JE,  Miao G;
PI   Rafalski JA,  Thorpe CJ;
XX
DR   WPI; 2000-182431/16.
DR   P-PSDB; AAY58837.
XX
PT   Nucleic acid fragments encoding cellulose biosynthetic enzyme useful as
PT   probes for isolating cDNAs and genes encoding homologous proteins, for
PT   producing transgenic plants
XX
PS   Claim 3; Page 55-56; 93pp; English.
XX
CC   The present sequence is that of full-length cDNA clone
CC   sdp2c.pk005.o22 encoding soybean cellulose synthase (CS, see AAY58837).
CC   The cDNA clone was isolated from a cDNA library produced from
CC   soybean pod tissue.  The invention relates to isolated nucleic acid
CC   fragments encoding plant CS and to CS polypeptides.  It also relates
CC   to the construction of a chimeric gene encoding all or a portion of
CC   the CS, in sense or antisense orientation, where expression of the
CC   gene results in altered levels of the CS in transformed host cells.
CC   The host cells can be used to screen compounds for their ability to
CC   inhibit CS activity.  CS nucleic acids are also useful for producing
CC   transgenic plants having altered levels of CS, and hence altered
CC   levels of fibre.  CS may also serve as a target for the development
CC   of novel herbicides.
XX
SQ   Sequence 3517 BP; 978 A; 686 C; 841 G; 1012 T; 0 other;

  Query Match             0.8%;  Score 27;  DB 21;  Length 3517;
  Best Local Similarity   100.0%;  Pred. No. 0.01;
  Matches   27;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2282 GGATATGAAGATAAAACTGAATGGGGA 2308
             ||||||||||||||||||||||||||||
Db       2310 GGATATGAAGATAAAACTGAATGGGGA 2336


RESULT 13
AAZ99500
ID   AAZ99500 standard; DNA; 3725 BP.
XX
AC   AAZ99500;
XX
DT   03-JUL-2000  (first entry)
XX
DE   DNA encoding a maize cellulose synthase
XX
KW   Maize; cellulose synthase; stalk quality; stand; silage; cellulose;
KW   transgenic plant; plant breeding marker; ss.
XX
OS   Zea mays.
XX
FH   Key             Location/Qualifiers
FT   CDS             179..3398
FT                   /*tag=  a
FT                   /product= "cellulose synthase"
FT                   /note= "no termination codon given"
XX
PN   WO200039706-A2.
XX
PD   24-FEB-2000.
XX
PF   16-AUG-1999;   99WO-US18760.
XX
PR   17-AUG-1998;   98US-0096822.
XX
PA   (PION-) PIONEER HI-BRED INT INC.
XX
PI   Dhugga KS,  Helentjaris TG,  Bowen BA,  Wang X;
XX
DR   WPI; 2000-224343/19.
DR   P-PSDB; AAY84110.
XX
PT   New genes which encode maize cellulose synthase polypeptides in plants
PT   useful for modulating the expression of cellulose synthase in plants
PT   and to produce transgenic plants expressing the novel protein
XX
PS   Claim 1; Page 105-110; 119pp; English.
XX
CC   The present sequence encodes a maize cellulose synthase polypeptide.
CC   The cellulose synthase can be used for the improvement of stalk quality
CC   for improved stand or silage. It also provides an increased concentration
CC   of cellulose in the pericarp, hardening the kernel and improving its
CC   handling ability. The sequences are used to produce transgenic plants
CC   and seeds expressing the cellulose synthase. The polynucleotide is
CC   used for modulating, preferably increasing, the level of the synthase
CC   in a plant cell. The plants are preferably monocots. The polynucleotide
CC   is also used as a probe or primer in the detection quantitation or
CC   isolation of gene transcripts. The probes are useful in detecting
CC   deficiencies in the level of mRNA in screenings for desired transgenic
CC   plant, for detecting mutations in the gene, for monitoring upregulation
CC   of expression or changes in enzyme activity in screening assays of
CC   compounds, for detection of any number of allelic variants of the gene,
CC   or for use as molecular markers in plant breeding programs. The
CC   isolated nucleic acids of the present invention can also be used for
CC   recombinant expression of their encoded polypeptides or for use as
CC   immunogens in the preparation and/or screening of antibodies. The
CC   proteins can be employed in assays for enzyme agonists or antagonists
CC   of enzyme function or for use of immunogens or antigens to obtain
CC   antibodies specifically immunoreactive with a protein.
XX
SQ   Sequence 3725 BP; 927 A; 807 C; 950 G; 1041 T; 0 other;

  Query Match             0.8%;  Score 26;  DB 21;  Length 3725;
  Best Local Similarity   100.0%;  Pred. No. 0.033;
  Matches   26;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1700 AGCAAAGCTCTTAGAGAAGCAATGTG 1725
             ||||||||||||||||||||||||||
Db       1868 AGCAAAGCTCTTAGAGAAGCAATGTG 1893


RESULT 14
AAZ99506
ID   AAZ99506 standard; DNA; 3725 BP.
XX
AC   AAZ99506;
XX
DT   03-JUL-2000  (first entry)
XX
DE   DNA encoding a maize cellulose synthase.
XX
KW   Maize; cellulose synthase; stalk quality; stand; silage; cellulose;
KW   transgenic plant; plant breeding marker; ss.
XX
OS   Zea mays.
XX
FH   Key             Location/Qualifiers
FT   CDS             179..3397
FT                   /*tag=  a
FT                   /product= "cellulose synthase"
FT                   /note= "no termination codon given"
XX
```

```
N   WO200009706-A2.
X
D   24-FEB-2000.
X
F   16-AUG-1999;   99WO-US18760.
X
R   17-AUG-1998;   98US-0096822.
X
A   (PION-) PIONEER HI-BRED INT INC.
X
I   Dhugga KS,  Helentjaris TG,  Bowen BA,  Wang X;
X
R   WPI; 2000-224343/19.
R   P-PSDB; AAY84112.
X
T   New genes which encode maize cellulose synthase polypeptides in plants
T   useful for modulating the expression of cellulose synthase in plants
T   and to produce transgenic plants expressing the novel protein
X
S   Claim 1; Page 121-126; 119pp; English.
X
C   The present sequence encodes a maize cellulose synthase polypeptide.
C   The cellulose synthase can be used for the improvement of stalk quality
C   for improved stand or silage. It also provides an increased concentration
C   of cellulose in the pericarp, hardening the kernel and improving its
C   handling ability. The sequences are used to produce transgenic plants
C   and seeds expressing the cellulose synthase. The polynucleotide is
C   used for modulating, preferably increasing, the level of the synthase
C   in a plant cell. The plants are preferably monocots. The polynucleotide
C   is also used as a probe or primer in the detection quantitation or
C   isolation of gene transcripts. The probes are useful in detecting
C   deficiencies in the level of mRNA in screenings for desired transgenic
C   plant, for detecting mutations in the gene, for monitoring upregulation
C   of expression or changes in enzyme activity in screening assays of
C   compounds, for detection of any number of allelic variants of the gene,
C   or for use as molecular markers in plant breeding programs. The
C   isolated nucleic acids of the present invention can also be used for
C   recombinant expression of their encoded polypeptides or for use as
C   immunogens in the preparation and/or screening of antibodies. The
C   proteins can be employed in assays for enzyme agonists or antagonists
C   of enzyme function or for use of immunogens or antigens to obtain
C   antibodies specifically immunoreactive with a protein.
X
Q   Sequence 3725 BP; 927 A; 807 C; 950 G; 1041 T; 0 other;

 Query Match             0.8%;  Score 26;  DB 21;  Length 3725;
 Best Local Similarity   100.0%;  Pred. No. 0.033;
 Matches   26;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

y       1700 AGCAAAGCTCTTAGAGAAGCAATGTG 1725
             ||||||||||||||||||||||||||
b       1868 AGCAAAGCTCTTAGAGAAGCAATGTG 1893


RESULT 15
AAZ99521
 D   AAZ99521 standard; DNA; 3725 BP.
XX
AC   AAZ99521;
XX
DT   03-JUL-2000  (first entry)
XX
DE   DNA encoding a maize cellulose synthase.
XX
KW   Maize; cellulose synthase; stalk quality; stand; silage; cellulose;
KW   transgenic plant; plant breeding marker; ss.
XX
OS   Zea mays.
XX
FH   Key             Location/Qualifiers
FT   CDS             179..3397
FT                   /*tag= a
FT                   /product= "cellulose synthase"
FT                   /note= "no termination codon given"
XX
PN   WO200009706-A2
XX
PD   24-FEB-2000.
XX
PF   16-AUG-1999;   99WO-US18760
XX
PR   17-AUG-1998;   98US-0096822
XX
PA   (PION-) PIONEER HI-BRED INT INC.
XX
PI   Dhugga KS,  Helentjaris TG,  Bowen BA,  Wang X;
XX
DR   WPI; 2000-224343/19.
DR   P-PSDB; AAY84112.
XX
PT   New genes which encode maize cellulose synthase polypeptides in plants
PT   useful for modulating the expression of cellulose synthase in plants
PT   and to produce transgenic plants expressing the novel protein
XX
PS   Claim 1; Page 121-126; 119pp; English.
XX
CC   The present sequence encodes a maize cellulose synthase polypeptide.
CC   The cellulose synthase can be used for the improvement of stalk quality
CC   for improved stand or silage. It also provides an increased concentration
CC   of cellulose in the pericarp, hardening the kernel and improving its
CC   handling ability. The sequences are used to produce transgenic plants
CC   and seeds expressing the cellulose synthase. The polynucleotide is
CC   used for modulating, preferably increasing, the level of the synthase
CC   in a plant cell. The plants are preferably monocots. The polynucleotide
CC   is also used as a probe or primer in the detection quantitation or
CC   isolation of gene transcripts. The probes are useful in detecting
CC   deficiencies in the level of mRNA in screenings for desired transgenic
CC   plant, for detecting mutations in the gene, for monitoring upregulation
CC   of expression or changes in enzyme activity in screening assays of
CC   compounds, for detection of any number of allelic variants of the gene,
CC   or for use as molecular markers in plant breeding programs. The
CC   isolated nucleic acids of the present invention can also be used for
CC   recombinant expression of their encoded polypeptides or for use as
CC   immunogens in the preparation and/or screening of antibodies. The
CC   proteins can be employed in assays for enzyme agonists or antagonists
CC   of enzyme function or for use of immunogens or antigens to obtain
CC   antibodies specifically immunoreactive with a protein.
XX
SQ   Sequence 3725 BP; 927 A; 807 C; 950 G; 1041 T; 0 other;

  Query Match             0.8%;  Score 26;  DB 21;  Length 3725;
  Best Local Similarity   100.0%;  Pred. No. 0.033;
  Matches   26;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1700 AGCAAAGCTCTTAGAGAAGCAATGTG 1725
             ||||||||||||||||||||||||||
Db      1868 AGCAAAGCTCTTAGAGAAGCAATGTG 1893


Search completed: October 30, 2003, 23:10:35
Job time : 840.27 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic nucleic search, using sw model

Run on: October 30, 2003, 22:15:30 ; Search time 182.305 Seconds
(without alignments)
8347.618 Million cell updates/sec

Title: US-09-980-043-5
Perfect score: 3444
Sequence: 1 gcggccgcggttaatgcgcg.....gtgttgtcgatagcggccgc 3444

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 569978 seqs, 220691866 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		32	0.9	3828	4	US-09-221-013A-7	Sequence 7, Appli
2		32	0.9	5009	4	US-09-221-013A-4	Sequence 4, Appli
3		29	0.8	3328	3	US-08-960-048-1	Sequence 1, Appli
4		29	0.8	3328	4	US-09-048-586-1	Sequence 1, Appli
5		21	0.6	3614	4	US-09-221-013A-9	Sequence 9, Appli
c	6	19	0.6	442	2	US-08-174-[illegible]-47	Sequence 47, Appl
c	7	19	0.6	442	2	US-08-498-[illegible]-47	Sequence 47, Appl
c	8	19	0.6	442	2	US-08-483-528B-36	Sequence 36, Appl
c	9	19	0.6	442	2	US-08-483-528B-53	Sequence 53, Appl
c	10	19	0.6	442	3	US-08-673-799C-36	Sequence 36, Appl
c	11	19	0.6	442	3	US-08-673-799C-53	Sequence 53, Appl
c	12	19	0.6	442	4	US-09-393-385B-36	Sequence 36, Appl
c	13	19	0.6	442	4	US-09-393-385B-53	Sequence 53, Appl
	14	19	0.6	686	4	US-09-166-350-9	Sequence 9, Appli
	15	19	0.6	1320	4	US-09-328-352-1936	Sequence 1936, Ap
	16	19	0.6	9751	3	US-09-248-303-7	Sequence 7, Appli
	17	19	0.6	9751	4	US-09-946-239-7	Sequence 7, Appli
c	18	19	0.6	1230025	4	US-09-198-452A-1	Sequence 1, Appli
	19	18	0.5	24	4	US-09-221-013A-18	Sequence 18, Appl
c	20	18	0.5	39	3	US-08-511-759B-7	Sequence 7, Appli
c	21	18	0.5	39	3	US-09-592-197-7	Sequence 7, Appli
c	22	18	0.5	39	4	US-08-892-724-7	Sequence 7, Appli
c	23	18	0.5	61	2	US-08-116-[illegible]-27	Sequence 27, Appl
c	24	18	0.5	61	2	US-08-498-562-27	Sequence 27, Appl
c	25	18	0.5	61	2	US-08-483-528B-27	Sequence 27, Appl
c	26	18	0.5	61	3	US-08-673-799C-27	Sequence 27, Appl
c	27	18	0.5	61	4	US-09-393-385B-27	Sequence 27, Appl
c	28	18	0.5	442	2	US-08-483-528B-52	Sequence 52, Appl
c	29	18	0.5	442	3	US-08-673-799C-52	Sequence 52, Appl
c	30	18	0.5	442	4	US-09-393-385B-52	Sequence 52, Appl
	31	18	0.5	1284	4	US-09-319-648-16	Sequence 16, Appl
	32	18	0.5	1578	3	US-09-044-404A-1	Sequence 1, Appli
	33	18	0.5	1578	4	US-09-586-924-1	Sequence 1, Appli
	34	18	0.5	2248	4	US-09-221-013A-1	Sequence 1, Appli
	35	18	0.5	3603	4	US-09-221-013A-5	Sequence 5, Appli
	36	18	0.5	3673	4	US-09-221-013A-11	Sequence 11, Appl
	37	18	0.5	5361	3	US-08-973-462-2	Sequence 2, Appli
c	38	18	0.5	5757	3	US-08-984-618-1	Sequence 1, Appli
	39	18	0.5	6152	3	US-08-973-462-1	Sequence 1, Appli
c	40	18	0.5	6256	4	US-09-620-312D-269	Sequence 269, App
	41	18	0.5	8411	4	US-09-221-013A-3	Sequence 3, Appli
	42	18	0.5	58407	4	US-08-916-421B-2	Sequence 2, Appli
	43	18	0.5	124884	4	US-09-661-596A-76	Sequence 76, Appl
c	44	18	0.5	124884	4	US-09-661-596A-76	Sequence 76, Appl
c	45	18	0.5	246240	2	US-08-724-394A-20	Sequence 20, Appl

ALIGNMENTS

RESULT 1
US-09-221-013A-7
; Sequence 7, Application US/09221013A
; Patent No. 6495740
; GENERAL INFORMATION:
; APPLICANT: Arioli, Antonio
; APPLICANT: Williamson, Richard E.
; APPLICANT: Betzner, Andreas S.
; APPLICANT: Peng, Liangcai
; TITLE OF INVENTION: Manipulation of cellulose and/or Beta-1,4-glucan
; FILE REFERENCE: 96-98
; CURRENT APPLICATION NUMBER: US/09/221,013A
; CURRENT FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: PCT/AU97/00402
; PRIOR FILING DATE: 1997-06-24
; PRIOR APPLICATION NUMBER: AU PO0699
; PRIOR FILING DATE: 1996-06-27
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 3828
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (239)..(3490)
US 09-221-013A 7

Query Match 0.9%; Score 32; DB 4; Length 3828;
Best Local Similarity 100.0%; Pred. No. 1.2e-05;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 419 TTGGATTCAGCTCCACCTGGCTCTCAGATTCC 450
||||||||||||||||||||||||||||||||
Db 653 TTGGATTCAGCTCCACCTGGCTCTCAGATTCC 684

RESULT 2
US-09-221-013A-4
; Sequence 4, Application US/09221013A
; Patent No. 6495740
; GENERAL INFORMATION:
; APPLICANT: Arioli, Antonio
; APPLICANT: Williamson, Richard E.
; APPLICANT: Betzner, Andreas S.
; APPLICANT: Peng, Liangcai
; TITLE OF INVENTION: Manipulation of cellulose and/or Beta-1,4-glucan
; FILE REFERENCE: 96-98
; CURRENT APPLICATION NUMBER: US/09/221,013A
; CURRENT FILING DATE: 1998-12-23

PRIOR APPLICATION NUMBER: PCT/AU97/00402
PRIOR FILING DATE: 1997-06-24
PRIOR APPLICATION NUMBER: AU PO0699
PRIOR FILING DATE: 1996-06-27
NUMBER OF SEQ ID NOS: 37
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 4
LENGTH: 5009
TYPE: DNA
ORGANISM: Arabidopsis thaliana
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(5009)
OTHER INFORMATION: N is A, T, G or C.
-09-221-013A-4

Query Match 0.9%; Score 32; DB 4; Length 5009;
Best Local Similarity 100.0%; Pred. No. 1.2e-05;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

/ 419 TTGGATTCAGCTCCACCTGGCTCTCAGATTCC 450
||||||||||||||||||||||||||||||||
b 1787 TTGGATTCAGCTCCACCTGGCTCTCAGATTCC 1818

ESULT 3
S-08-960-048-1
Sequence 1, Application US/08960048C
Patent No. 6271443
GENERAL INFORMATION:
APPLICANT: Stalker, D. et al.
TITLE OF INVENTION: Plant Cellulose Synthase and Promoter
TITLE OF INVENTION: Sequences
FILE REFERENCE: 15621/01/US
CURRENT APPLICATION NUMBER: US/08/960,048C
CURRENT FILING DATE: 1997-10-29
PRIOR APPLICATION NUMBER: 60/029,987
PRIOR FILING DATE: 1996-10-29
NUMBER OF SEQ ID NOS: 12
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 1
LENGTH: 3328
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthetic Oligonucleotide
S-08-960-048-1

Query Match 0.8%; Score 29; DB 3; Length 3328;
Best Local Similarity 100.0%; Pred. No. 0.00035;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

y 1070 AGTACAGTGGATCCATTGAAAGAGCCTCC 1098
|||||||||||||||||||||||||||||
b 947 AGTACAGTGGATCCATTGAAAGAGCCTCC 975

ESULT 4
S-09-838-586-1
Sequence 1, Application US/09838586
Patent No. 6576818
GENERAL INFORMATION:
APPLICANT: Stalker, D. et al.
TITLE OF INVENTION: Plant Cellulose Synthase and Promoter
TITLE OF INVENTION: Sequences
FILE REFERENCE: 15621/02/US
CURRENT APPLICATION NUMBER: US/09/838,586
CURRENT FILING DATE: 2001-04-18
PRIOR APPLICATION NUMBER: 60/029,987
PRIOR FILING DATE: 1996-10-29
PRIOR APPLICATION NUMBER: 08/960,048
PRIOR FILING DATE: 1997-10-29
NUMBER OF SEQ ID NOS: 12

; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 3328
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Oligonucleotide
US-09-838-586-1

Query Match 0.8%; Score 29; DB 4; Length 3328;
Best Local Similarity 100.0%; Pred. No. 0.00035;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1070 AGTACAGTGGATCCATTGAAAGAGCCTCC 1098
|||||||||||||||||||||||||||||
Db 947 AGTACAGTGGATCCATTGAAAGAGCCTCC 975

RESULT 5
US-09-221-013A-9
; Sequence 9, Application US/09221013A
; Patent No. 6495740
; GENERAL INFORMATION:
; APPLICANT: Arioli, Antonio
; APPLICANT: Williamson, Richard E
; APPLICANT: Betzner, Andreas S.
; APPLICANT: Peng, Liangc
; TITLE OF INVENTION: Manipulation of cellulose and/or Beta 1,4 glucan
; FILE REFERENCE: 96-98
; CURRENT APPLICATION NUMBER: US/09/221,013A
; CURRENT FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: PCT/AU97/00402
; PRIOR FILING DATE: 1997-06-24
; PRIOR APPLICATION NUMBER: AU PO0699
; PRIOR FILING DATE: 1996-06-27
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 3614
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (217)..(3411)
US-09-221-013A-9

Query Match 0.6%; Score 21; DB 4; Length 3614;
Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1172 TCTGATGATGGTGCTGCTATG 1192
|||||||||||||||||||||
Db 1345 TCTGATGATGGTGCTGCTATG 1365

RESULT 6
US-08-116-778E-47/c
; Sequence 47, Application US/08116778E
; Patent No. 5830470
; GENERAL INFORMATION:
; APPLICANT: NAKAMURA, KAZUYASU
; APPLICANT: KOIKE, MASAMICHI
; APPLICANT: SHITARA, KENYA
; APPLICANT: HANAI, NOBUO
; APPLICANT: KUWANA, YOSHIHISA
; APPLICANT: HASEGAWA, MAMORU
; TITLE OF INVENTION: HUMANIZED ANTIBODIES
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA

```
;       COUNTRY:  U.S.A.
;       ZIP:  22201-4714
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/116,778E
;       FILING DATE:  07-SEP-93
;       CLASSIFICATION:  424
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  WILSON, MARY J.
;       REGISTRATION NUMBER:  32,955
;       REFERENCE/DOCKET NUMBER:  249-59
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (703)816-4000
;       TELEFAX:  (703)816-4100
;   INFORMATION FOR SEQ ID NO:  47:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  442 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  both
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  other nucleic acid
US-08-116-778E-47

  Query Match             0.6%;  Score 19;  DB 2;  Length 442;
  Best Local Similarity   100.0%;  Pred. No. 25;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2842 GGTTCTTGCTGGTGTCGAC 2860
             ||||||||||||||||||||
Db        298 GGTTCTTGCTGGTGTCGAC 280


RESULT 7
US-08-438-562-47/c
; Sequence 47, Application US/08438562
; Patent No. 5874255
;  GENERAL INFORMATION:
;    APPLICANT:  NAKAMURA, KAZUYASU
;    APPLICANT:  KOIKE, MASAMICHI
;    APPLICANT:  SHITARA, KENYA
;    APPLICANT:  HANAI, NOBUO
;    APPLICANT:  KUWANA, YOSHIHISA
;    APPLICANT:  HASEGAWA, MAMORU
;    TITLE OF INVENTION:  HUMANIZED ANTIBODIES
;    NUMBER OF SEQUENCES:  49
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NIXON & VANDERHYE P.C.
;      STREET:  1100 NORTH GLEBE ROAD
;      CITY:  ARLINGTON
;      STATE:  VIRGINIA
;      COUNTRY:  U.S.A.
;      ZIP:  22201-4714
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/438,562
;      FILING DATE:  10-MAY-95
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/116,778
;      FILING DATE:  07-SEP-93
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  WILSON, MARY J.
;      REGISTRATION NUMBER:  32,955
;      REFERENCE/DOCKET NUMBER:  249-76
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)816-4000
;      TELEFAX:  (703)816-4100
;  INFORMATION FOR SEQ ID NO:  47:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  442 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  both
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
US-08-438-562-47

  Query Match             0.6%;  Score 19;  DB 2;  Length 442;
  Best Local Similarity   100.0%;  Pred. No. 25;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2842 GGTTCTTGCTGGTGTCGAC 2860
             ||||||||||||||||||||
Db        298 GGTTCTTGCTGGTGTCGAC 280


RESULT 8
US-08-483-528B-36/c
; Sequence 36, Application US/08483528B
; Patent No. 5939532
;  GENERAL INFORMATION:
;    APPLICANT:  NAKAMURA, KAZUYASU
;    APPLICANT:  KOIKE, MASAMICHI
;    APPLICANT:  SHITARA, KENYA
;    APPLICANT:  HANAI, NOBUO
;    APPLICANT:  KUWANA, YOSHIHISA
;    APPLICANT:  HASEGAWA, MAMORU
;    TITLE OF INVENTION:  HUMANIZED ANTIBODIES
;    NUMBER OF SEQUENCES:  103
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NIXON & VANDERHYE P.C.
;      STREET:  1100 NORTH GLEBE ROAD
;      CITY:  ARLINGTON
;      STATE:  VIRGINIA
;      COUNTRY:  U.S.A.
;      ZIP:  22201-4714
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/483,528B
;      FILING DATE:  07-JUN-95
;      CLASSIFICATION:  536
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)816-4000
;      TELEFAX:  (703)816-4100
;  INFORMATION FOR SEQ ID NO:  36:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  442 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-483-528B-36

  Query Match             0.6%;  Score 19;  DB 2;  Length 442;
  Best Local Similarity   100.0%;  Pred. No. 25;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2842 GGTTCTTGCTGGTGTCGAC 2860
             ||||||||||||||||||||
Db        298 GGTTCTTGCTGGTGTCGAC 280


RESULT 9
US-08-483-528B-53/c
```

```
Sequence 53, Application US/08483528B
Patent No. 5939532
 GENERAL INFORMATION:
   APPLICANT:  NAKAMURA, KAZUYASU
   APPLICANT:  KOIKE, MASAMICHI
   APPLICANT:  SHITARA, KENYA
   APPLICANT:  HANAI, NOBUO
   APPLICANT:  KUWANA, YOSHIHISA
   APPLICANT:  HASEGAWA, MAMORU
   TITLE OF INVENTION:  HUMANIZED ANTIBODIES
   NUMBER OF SEQUENCES:  103
   CORRESPONDENCE ADDRESS:
     ADDRESSEE:  NIXON & VANDERHYE P.C.
     STREET:  1100 NORTH GLEBE ROAD
     CITY:  ARLINGTON
     STATE:  VIRGINIA
     COUNTRY:  U.S.A.
     ZIP:  22201-4714
   COMPUTER READABLE FORM:
     MEDIUM TYPE:  Floppy disk
     COMPUTER:  IBM PC compatible
     OPERATING SYSTEM:  PC-DOS/MS-DOS
     SOFTWARE:  PatentIn Release #1.0, Version #1.25
   CURRENT APPLICATION DATA:
     APPLICATION NUMBER:  US/08/483,528B
     FILING DATE:  07-JUN-95
     CLASSIFICATION:  536
   TELECOMMUNICATION INFORMATION:
     TELEPHONE:  (703)816-4000
     TELEFAX:  (703)816-4100
 INFORMATION FOR SEQ ID NO:  53:
   SEQUENCE CHARACTERISTICS:
     LENGTH:  442 base pairs
     TYPE:  nucleic acid
     STRANDEDNESS:  double
     TOPOLOGY:  linear
   MOLECULE TYPE:  other nucleic acid
     DESCRIPTION:  /desc = "synthetic DNA"
   ORIGINAL SOURCE:
     ORGANISM:  Homo sapiens and mouse
   FEATURE:
     NAME/KEY:  sig_peptide
     LOCATION:  -19..-1
     IDENTIFICATION METHOD:  by similarity with known sequence or to an established co
   FEATURE:
     NAME/KEY:  domain
     LOCATION:  31..35
     IDENTIFICATION METHOD:  by similarity with known sequence or to an established co
     OTHER INFORMATION:  /product= "CDR1"
   FEATURE:
     NAME/KEY:  domain
     LOCATION:  50..66
     IDENTIFICATION METHOD:  by similarity with known sequence or to an established co
     OTHER INFORMATION:  /product= "CDR2"
   FEATURE:
     NAME/KEY:  domain
     LOCATION:  99..109
     IDENTIFICATION METHOD:  by similarity with known sequence or to an established co
     OTHER INFORMATION:  /product= "CDR3"
S-08-483-528B-53

 Query Match              0.6%;  Score 19;  DB 2;  Length 442;
 Best Local Similarity   100.0%;  Pred. No. 25;
 Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2842 GGTTCTTGCTGGTGTCGAC 2860
             ||||||||||||||||||||
Db        298 GGTTCTTGCTGGTGTCGAC 280


ESULT 10
JS-08-673-799C-36/c
 Sequence 36, Application US/08673799C
; Patent No. 6042828
;  GENERAL INFORMATION:
;    APPLICANT:  NAKAMURA, KAZUYASU
;    APPLICANT:  KOIKE, MASAMICHI
;    APPLICANT:  SHITARA, KENYA
;    APPLICANT:  HANAI, NOBUO
;    APPLICANT:  KUWANA, YOSHIHISA
;    APPLICANT:  HASEGAWA, MAMORU
;    TITLE OF INVENTION:  HUMANIZED ANTIBODIES
;    NUMBER OF SEQUENCES:  103
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NIXON & VANDERHYE P.C.
;      STREET:  1100 NORTH GLEBE ROAD
;      CITY:  ARLINGTON
;      STATE:  VIRGINIA
;      COUNTRY:  U.S.A.
;      ZIP:  22201-4714
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/673,799C
;      FILING DATE:  27-JUN-96
;      CLASSIFICATION:  536
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)816-4000
;      TELEFAX:  (703)816-4100
;  INFORMATION FOR SEQ ID NO:  36:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  442 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  both
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Other nucleic acid
US-08-673-799C-36

  Query Match              0.6%;  Score 19;  DB 3;  Length 442;
  Best Local Similarity   100.0%;  Pred. No. 25;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2842 GGTTCTTGCTGGTGTCGAC 2860
             ||||||||||||||||||||
Db        298 GGTTCTTGCTGGTGTCGAC 280


RESULT 11
US-08-673-799C-53/c
; Sequence 53, Application US/08673799C
; Patent No. 6042828
;  GENERAL INFORMATION:
;    APPLICANT:  NAKAMURA, KAZUYASU
;    APPLICANT:  KOIKE, MASAMICHI
;    APPLICANT:  SHITARA, KENYA
;    APPLICANT:  HANAI, NOBUO
;    APPLICANT:  KUWANA, YOSHIHISA
;    APPLICANT:  HASEGAWA, MAMORU
;    TITLE OF INVENTION:  HUMANIZED ANTIBODIES
;    NUMBER OF SEQUENCES:  103
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NIXON & VANDERHYE P.C.
;      STREET:  1100 NORTH GLEBE ROAD
;      CITY:  ARLINGTON
;      STATE:  VIRGINIA
;      COUNTRY:  U.S.A.
;      ZIP:  22201-4714
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
```

```
;       APPLICATION NUMBER:  US/08/673,799C
;       FILING DATE:  27-JUN-96
;       CLASSIFICATION:  536
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (703)816-4000
;       TELEFAX:  (703)816-4100
;   INFORMATION FOR SEQ ID NO:  53:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  442 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  other nucleic acid
;       DESCRIPTION:  /desc = "synthetic DNA"
;     ORIGINAL SOURCE:
;       ORGANISM:  Homo sapiens and mouse
;     FEATURE:
;       NAME/KEY:  sig_peptide
;       LOCATION:  -19..-1
;       IDENTIFICATION METHOD:  by similarity with known sequence or to an established cc
;     FEATURE:
;       NAME/KEY:  domain
;       LOCATION:  31..35
;       IDENTIFICATION METHOD:  by similarity with known sequence or to an established cc
;       OTHER INFORMATION:  /product= "CDR1"
;     FEATURE:
;       NAME/KEY:  domain
;       LOCATION:  50..66
;       IDENTIFICATION METHOD:  by similarity with known sequence or to an established cc
;       OTHER INFORMATION:  /product= "CDR2"
;     FEATURE:
;       NAME/KEY:  domain
;       LOCATION:  99..109
;       IDENTIFICATION METHOD:  by similarity with known sequence or to an established cc
;       OTHER INFORMATION:  /product= "CDR3"
US-08-673-799C-53

  Query Match             0.6%;  Score 19;  DB 3;  Length 442;
  Best Local Similarity   100.0%;  Pred. No. 25;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2842 GGTTCTTGCTGGTGTCGAC 2860
             ||||||||||||||||||||
Db        298 GGTTCTTGCTGGTGTCGAC 280


RESULT 12
US-09-393-385B-36/c
; Sequence 36, Application US/09393385B
; Patent No. 6423511
;  GENERAL INFORMATION:
;     APPLICANT:  NAKAMURA, KAZUYASU
;     APPLICANT:  KOIKE, MASAMICHI
;     APPLICANT:  SHITARA, KENYA
;     APPLICANT:  HANAI, NOBUO
;     APPLICANT:  KUWANA, YOSHIHISA
;     APPLICANT:  HASEGAWA, MAMORU
;     TITLE OF INVENTION:  HUMANIZED ANTIBODIES
;     NUMBER OF SEQUENCES:  113
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  NIXON & VANDERHYE P.C
;       STREET:  1100 NORTH GLEBE ROAD
;       CITY:  ARLINGTON
;       STATE:  VIRGINIA
;       COUNTRY:  U.S.A.
;       ZIP:  22201-4714
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/393,385B
;       FILING DATE:  27-JUN-96
;       CLASSIFICATION:
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (703)816-4000
;       TELEFAX:  (703)816-4100
;   INFORMATION FOR SEQ ID NO:  36:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  442 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  both
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  Other nucleic acid
US-09-393-385B-36

  Query Match             0.6%;  Score 19;  DB 4;  Length 442;
  Best Local Similarity   100.0%;  Pred. No. 25;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2842 GGTTCTTGCTGGTGTCGAC 2860
             ||||||||||||||||||||
Db        298 GGTTCTTGCTGGTGTCGAC 280


RESULT 13
US-09-393-385B-53/c
; Sequence 53, Application US/09393385B
; Patent No. 6423511
; GENERAL INFORMATION:
;   APPLICANT:  NAKAMURA, KAZUYASU
;   APPLICANT:  KOIKE, MASAMICHI
;   APPLICANT:  SHITARA, KENYA
;   APPLICANT:  HANAI, NOBUO
;   APPLICANT:  KUWANA, YOSHIHISA
;   APPLICANT:  HASEGAWA, MAMORU
;   TITLE OF INVENTION:  HUMANIZED ANTIBODIES
;   NUMBER OF SEQUENCES:  113
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  NIXON & VANDERHYE P.C.
;     STREET:  1100 NORTH GLEBE ROAD
;     CITY:  ARLINGTON
;     STATE:  VIRGINIA
;     COUNTRY:  U.S.A.
;     ZIP:  22201-4714
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/09/393,385B
;     FILING DATE:  27-JUN-96
;     CLASSIFICATION:
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  (703)816-4000
;     TELEFAX:  (703)816-4100
; INFORMATION FOR SEQ ID NO:  53:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  442 base pairs
;     TYPE:  nucleic acid
;     STRANDEDNESS:  double
;     TOPOLOGY:  linear
;   MOLECULE TYPE:  other nucleic acid
;     DESCRIPTION:  /desc = "synthetic DNA"
;   ORIGINAL SOURCE:
;     ORGANISM:  Homo sapiens and mouse
;   FEATURE:
;     NAME/KEY:  sig_peptide
;     LOCATION:  10..66
;     IDENTIFICATION METHOD:  by similarity with known sequence or to an
;     IDENTIFICATION METHOD:  established consensus
;   FEATURE:
;     NAME/KEY:  domain
;     LOCATION:  157..171
```

IDENTIFICATION METHOD: by similarity with known sequence or to an
IDENTIFICATION METHOD: established consensus
OTHER INFORMATION: /product= "CDR1"
FEATURE:
NAME/KEY: domain
LOCATION: 214..264
IDENTIFICATION METHOD: by similarity with known sequence or to an
IDENTIFICATION METHOD: established consensus
OTHER INFORMATION: /product= "CDR2"
FEATURE:
NAME/KEY: domain
LOCATION: 361..393
IDENTIFICATION METHOD: by similarity with known sequence or to an
IDENTIFICATION METHOD: established consensus
OTHER INFORMATION: /product= "CDR3"
S-09-393-385B-53

Query Match 0.6%; Score 19; DB 4; Length 442;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2842 GGTTCTTGCTGGTGTCGAC 2860
||||||||||||||||||
Db 298 GGTTCTTGCTGGTGTCGAC 280

RESULT 14
S-09-166-350-9
Sequence 9, Application US/09166350A
Patent No. 6440663
GENERAL INFORMATION:
APPLICANT: Scanlan, Matthew
APPLICANT: Chen, Yao
APPLICANT: Stockert, Elisabeth
APPLICANT: Old, Lloyd
APPLICANT: Jager, Elke
APPLICANT: Knuth, Alex
TITLE OF INVENTION: Renal Cancer Associated Antigens and
TITLE OF INVENTION: Uses Therefor
FILE REFERENCE: L0461/7051
CURRENT APPLICATION NUMBER: US/09/166,350A
CURRENT FILING DATE: 1998-10-05
EARLIER APPLICATION NUMBER: US 09/166,350
EARLIER FILING DATE: 1998-10-05
NUMBER OF SEQ ID NOS: 35
SOFTWARE: FastSEQ for Windows Version 3.0
SEQ ID NO 9
LENGTH: 686
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: unsure
LOCATION: 43..43
OTHER INFORMATION: n = a, c, g or t
FEATURE:
NAME/KEY: unsure
LOCATION: 64..64
OTHER INFORMATION: n = a, c, g or t
FEATURE:
NAME/KEY: unsure
LOCATION: 284..284
OTHER INFORMATION: n = a, c, g or t
FEATURE:
NAME/KEY: unsure
LOCATION: 404..404
OTHER INFORMATION: n = a, c, g or t
FEATURE:
NAME/KEY: unsure
LOCATION: 624..624
OTHER INFORMATION: n = a, c, g or t
S-09-166-350-9

Query Match 0.6%; Score 19; DB 4; Length 686;
Best Local Similarity [illegible]; Pred. No. 26;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 296 GAGGAAGATGACAT[illegible]ATT 314
Db 218 GAGGAAGATGACAT[illegible]ACA 236

RESULT 19
US-09-328-352-1936
; Sequence 1936, Application US/09328452
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: [illegible]
; CURRENT APPLICATION NUMBER: US/09/328,452
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 4[illegible]
; SEQ ID NO 1936
; LENGTH: 1320
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-1936

Query Match 0.6%; Score 19; DB 4; Length 1320;
Best Local Similarity 1[illegible].0%; Pred. No. 34;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 522 TTGGTGGTCATGG[illegible]ATAG 540
Db 665 TTGGTGGTCATGG[illegible]ATAG 683

Search completed: October 31, 2003, 09:33:01
Job time : 184.105 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 31, 2003, 05:47:13 ; Search time 857.389 Seconds
(without alignments)
10924.991 Million cell updates/sec

Title: US-09-980-043-5
Perfect score: 3444
Sequence: 1 gcggccgcggttaatcgccg .....gtgttgtgataggcggccgc 3444

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1811591 seqs, 1359896290 residues

Word size: 0

Total number of hits satisfying chosen parameters: 3623182

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post processing: Listing first 45 summaries

Database: Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the actual database distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2322	67.4	3255	10	US-09-938-842A-2530	Sequence 2530, Ap
c 2	361	10.5	447	9	US-09-770-444-763	Sequence 763, App
c 3	222	6.4	342	10	US-09-924-035A-676	Sequence 676, App
4	32	0.9	2890	10	US-09-900-237-17	Sequence 17, Appl
5	32	0.9	3255	10	US-09-938-842A-559	Sequence 559, App
6	32	0.9	3828	14	US-10-229-193-7	Sequence 7, Appli
7	32	0.9	5009	14	US-10-229-193-4	Sequence 4, Appli
c 8	29	0.8	489	11	US-09-770-961-707	Sequence 707, App
9	29	0.8	3231	12	US-10-209-059-27	Sequence 27, Appl
10	29	0.8	3328	10	US-09-838-539-1	Sequence 1, Appli
11	29	0.8	3704	12	US-10-209-059-13	Sequence 13, Appl
12	29	0.8	3704	12	US-10-160-719-57	Sequence 57, Appl
13	27	0.8	3517	10	US-09-900-237-13	Sequence 13, Appl
14	26	0.8	3725	12	US-10-209-059-45	Sequence 45, Appl
15	26	0.8	3725	12	US-10-160-719-13	Sequence 13, Appl
16	26	0.8	3725	12	US-10-160-719-21	Sequence 21, Appl
17	26	0.8	3725	12	US-10-160-719-41	Sequence 41, Appl
18	26	0.8	3753	12	US-10-160-719-33	Sequence 33, Appl
19	26	0.8	3753	12	US-10-160-719-53	Sequence 53, Appl
20	26	0.8	3780	12	US-10-209-059-1	Sequence 1, Appli
21	26	0.8	3780	12	US-10-160-719-9	Sequence 9, Appli
22	26	0.8	3786	10	US-09-900-237-7	Sequence 7, Appli
23	26	0.8	3799	12	US-10-267-459-5	Sequence 5, Appli
24	26	0.8	3799	12	US-10-209-059-21	Sequence 21, Appl
25	26	0.8	3813	12	US-10-209-059-17	Sequence 17, Appl
26	26	0.8	3813	12	US-10-160-719-25	Sequence 25, Appl
27	26	0.8	3813	12	US-10-160-719-45	Sequence 45, Appl
28	23	0.7	360	9	US-09-770-791-520	Sequence 520, App
29	23	0.7	2268	10	US-09-887-576-669	Sequence 669, App
30	23	0.7	2663	10	US-09-900-237-23	Sequence 23, Appl
31	23	0.7	3626	10	US-09-900-237-29	Sequence 29, Appl
32	23	0.7	3936	10	US-09-900-237-9	Sequence 9, Appli
33	23	0.7	3969	12	US-10-209-059-49	Sequence 49, Appl
34	23	0.7	3969	12	US-10-160-719-17	Sequence 17, Appl
35	23	0.7	3969	12	US-10-160-719-37	Sequence 37, Appl
36	22	0.6	439	10	US-09-983-965-301	Sequence 301, App
37	22	0.6	440	10	US-09-983-965-486	Sequence 486, App
38	22	0.6	2199	14	US-10-198-846-10002	Sequence 10002, A
39	22	0.6	2316	14	US-10-103-313-49	Sequence 49, Appl
40	22	0.6	2487	10	US-09-925-300-404	Sequence 404, App
41	22	0.6	3563	10	US-09-900-237-25	Sequence 25, Appl
42	21	0.6	327	12	US-09-814-353-3866	Sequence 3866, Ap
43	21	0.6	327	12	US-09-814-353-10175	Sequence 10175, A
44	21	0.6	403	12	US-09-814-353-16559	Sequence 16559, A
45	21	0.6	1560	10	US-09-900-237-27	Sequence 27, Appl

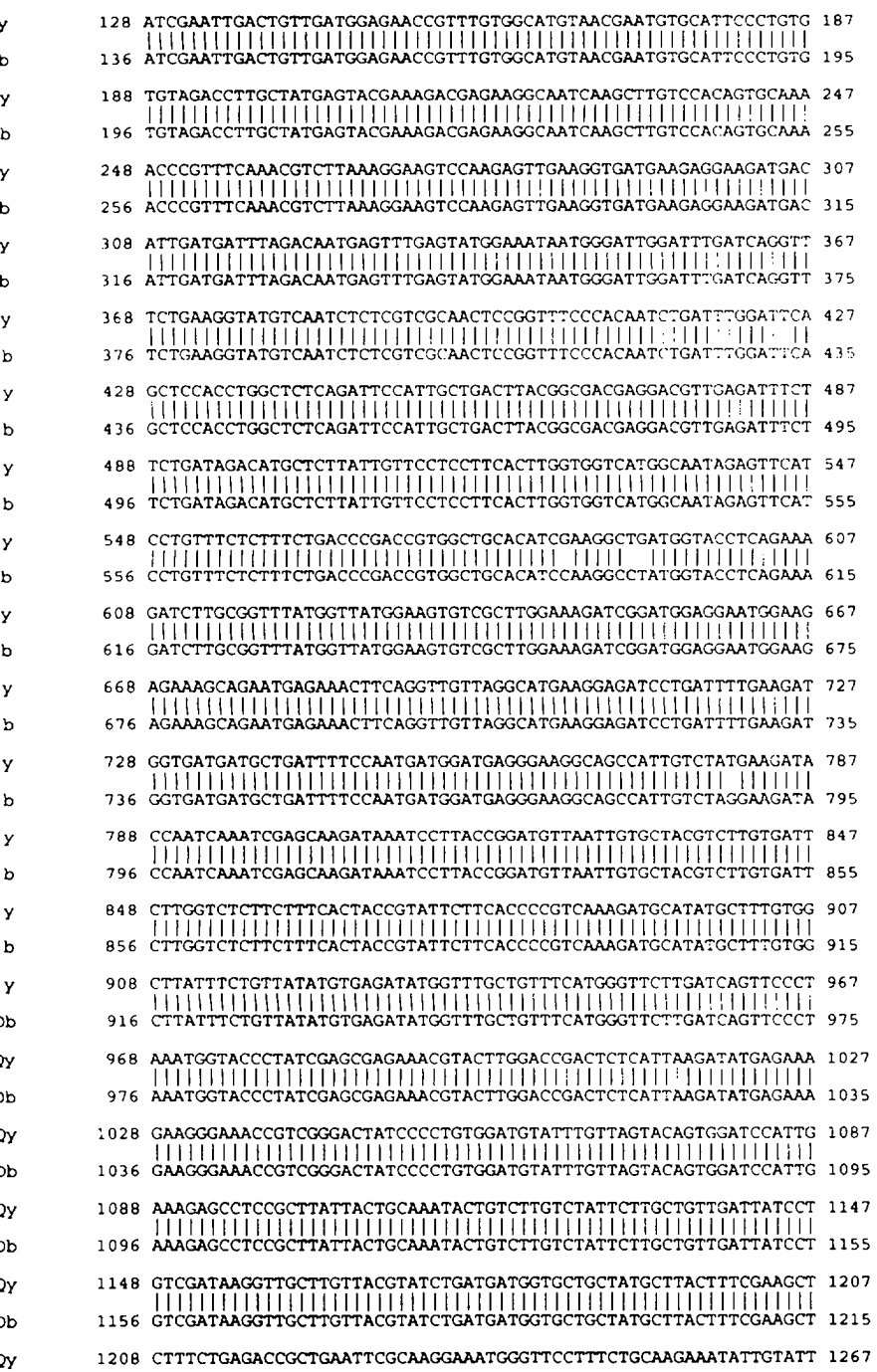
ALIGNMENTS

RESULT 1
US-09-938-842A-2530
; Sequence 2530, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRIP1300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 2530
; LENGTH: 3255
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-2530

Query Match 67.4%; Score 2322; DB 10; Length 3255;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 3222; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

```
Qy          8 CGGTTAATCGCCGGTTCTCACAACAGGAATGAGTTTGTCCTCATTAATGCCGATGAGAAT 67
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         16 CGGTTAATCGCCGGTTCTCACAACAGGAATGAGTTTGTCCTCATTAATGCCGATGAGAAT 75

Qy         68 GCCCGAATAAGATCAGTCCAAGAGCTGAGTGGACAGACATGTCAAATCTGCAGAGATGAG 127
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         76 GCCCGAATAAGATCAGTCCAAGAGCTGAGTGGACAGACATGTCAAATCTGCAGAGATGAG 135
```


Qy 128 ATCGAATTGACTGTTGATGGAGAACCGTTTGTGGCATGTAACGAATGTGCATTCCCTGTG 187
Db 136 ATCGAATTGACTGTTGATGGAGAACCGTTTGTGGCATGTAACGAATGTGCATTCCCTGTG 195
Qy 188 TGTAGACCTTGCTATGAGTACGAAAGACGAGAAGGCAATCAAGCTTGTCCACAGTGCAAA 247
Db 196 TGTAGACCTTGCTATGAGTACGAAAGACGAGAAGGCAATCAAGCTTGTCCACAGTGCAAA 255
Qy 248 ACCCGTTTCAAACGTCTTAAAGGAAGTCCAAGAGTTGAAGGTGATGAAGAGGAAGATGAC 307
Db 256 ACCCGTTTCAAACGTCTTAAAGGAAGTCCAAGAGTTGAAGGTGATGAAGAGGAAGATGAC 315
Qy 308 ATTGATGATTTAGACAATGAGTTTGAGTATGGAAATAATGGGATTGGATTTGATCAGGTT 367
Db 316 ATTGATGATTTAGACAATGAGTTTGAGTATGGAAATAATGGGATTGGATTTGATCAGGTT 375
Qy 368 TCTGAAGGTATGTCAATCTCTCGTCGCAACTCCGGTTTCCCACAATCTGATTTGGATTCA 427
Db 376 TCTGAAGGTATGTCAATCTCTCGTCGCAACTCCGGTTTCCCACAATCTGATTTGGATTCA 435
Qy 428 GCTCCACCTGGCTCTCAGATTCCATTGCTGACTTACGGCGACGAGGACGTTGAGATTTCT 487
Db 436 GCTCCACCTGGCTCTCAGATTCCATTGCTGACTTACGGCGACGAGGACGTTGAGATTTCT 495
Qy 488 TCTGATAGACATGCTCTTATTGTTCCTCCTTCACTTGGTGGTCATGGCAATAGAGTTCAT 547
Db 496 TCTGATAGACATGCTCTTATTGTTCCTCCTTCACTTGGTGGTCATGGCAATAGAGTTCAT 555
Qy 548 CCTGTTTCTCTTTCTGACCCGACCGTGGCTGCACATCGAAGGCTGATGGTACCTCAGAAA 607
Db 556 CCTGTTTCTCTTTCTGACCCGACCGTGGCTGCACATCCAAGGCCTATGGTACCTCAGAAA 615
Qy 608 GATCTTGCGGTTTATGGTTATGGAAGTGTCGCTTGGAAAGATCGGATGGAGGAATGGAAG 667
Db 616 GATCTTGCGGTTTATGGTTATGGAAGTGTCGCTTGGAAAGATCGGATGGAGGAATGGAAG 675
Qy 668 AGAAAGCAGAATGAGAAACTTCAGGTTGTTAGGCATGAAGGAGATCCTGATTTTGAAGAT 727
Db 676 AGAAAGCAGAATGAGAAACTTCAGGTTGTTAGGCATGAAGGAGATCCTGATTTTGAAGAT 735
Qy 728 GGTGATGATGCTGATTTTCCAATGATGGATGAGGGAAGGCAGCCATTGTCTATGAAGATA 787
Db 736 GGTGATGATGCTGATTTTCCAATGATGGATGAGGGAAGGCAGCCATTGTCTAGGAAGATA 795
Qy 788 CCAATCAAATCGAGCAAGATAAATCCTTACCGGATGTTAATTGTGCTACGTCTTGTGATT 847
Db 796 CCAATCAAATCGAGCAAGATAAATCCTTACCGGATGTTAATTGTGCTACGTCTTGTGATT 855
Qy 848 CTTGGTCTCTTCTTTCACTACCGTATTCTTCACCCCGTCAAAGATGCATATGCTTTGTGG 907
Db 856 CTTGGTCTCTTCTTTCACTACCGTATTCTTCACCCCGTCAAAGATGCATATGCTTTGTGG 915
Qy 908 CTTATTTCTGTTATATGTGAGATATGGTTTGCTGTTTCATGGGTTCTTGATCAGTTCCCT 967
Db 916 CTTATTTCTGTTATATGTGAGATATGGTTTGCTGTTTCATGGGTTCTTGATCAGTTCCCT 975
Qy 968 AAATGGTACCCTATCGAGCGAGAAACGTACTTGGACCGACTCTCATTAAGATATGAGAAA 1027
Db 976 AAATGGTACCCTATCGAGCGAGAAACGTACTTGGACCGACTCTCATTAAGATATGAGAAA 1035
Qy 1028 GAAGGGAAACCGTCGGGACTATCCCCTGTGGATGTATTTGTTAGTACAGTGGATCCATTG 1087
Db 1036 GAAGGGAAACCGTCGGGACTATCCCCTGTGGATGTATTTGTTAGTACAGTGGATCCATTG 1095
Qy 1088 AAAGAGCCTCCGCTTATTACTGCAAATACTGTCTTGTCTATTCTTGCTGTTGATTATCCT 1147
Db 1096 AAAGAGCCTCCGCTTATTACTGCAAATACTGTCTTGTCTATTCTTGCTGTTGATTATCCT 1155
Qy 1148 GTCGATAAGGTTGCTTGTTACGTATCTGATGATGGTGCTGCTATGCTTACTTTCGAAGCT 1207
Db 1156 GTCGATAAGGTTGCTTGTTACGTATCTGATGATGGTGCTGCTATGCTTACTTTCGAAGCT 1215
Qy 1208 CTTTCTGAGACCGCTGAATTCGCAAGGAAAATGGGTTCCTTTCTGCAAGAAATATTGTATT 1267
Db 1216 CTTTCTGAGACCGCTGAATTCGCAAGGAAAATGGGTTCCTTTCTGCAAGAAATATTGTATT 1275
Qy 1268 GAGCCTCGTGCTCG[illegible]GAATGGTATTTCTGCCATAAAATGGACTACTTGAAGAATAAAGTT 1327
Db 1276 GAGCCTCGTGCTCG[illegible]GAATGGTATTTCTGCCATAAAATGGACTACTTGAAGAATAAAGTT 1335
Qy 1328 CATCCCGCATTTGT[illegible]AGGGAGCGGCGAGCCATGAAGAGAGATTATGAAGAATTCAAAGTA 1387
Db 1336 CATCCCGCATTTGT[illegible]AGGGAGCGGCGAGCCATGAAGAGAGATTATGAAGAATTCAAAGTA 1395
Qy 1388 AAGATCAATGCTT[illegible]AGTAGCAACAGCACAGAAAGTGCCTGAGGATGGTTGGACTATGCAA 1447
Db 1396 AAGATCAATGCTT[illegible]AGTAGCAACAGCACAGAAAGTGCCTGAGGATGGTTGGACTATGCAA 1455
Qy 1448 GACGGTACAG[illegible]GAATA[illegible]ATCATCCTGGCATGATTCAGGTCTTC 1507
Db 1456 GACGGTACAG[illegible]GAATA[illegible]ATCATCCTGGCATGATTCAGGTCTTC 1515
Qy 1508 CTTGGAAGTGACT[illegible]GATTCAGAAAGAACGAGTTGCCTCGATTAGTTTACGTT 1567
Db 1516 CTTGGAAGTGACG[illegible]GATTCGTGATGTCGAAAACAACGAGTTGCCTCGATTAGTTTACGTT 1575
Qy 1568 TCTCGTGAGAAGACACCCGGATTTGATCACCATAAGAAGGCTGGAGCTATGAATTCCCTG 1627
Db 1576 TCTCGTGAGAAGACACCCGGATTTGATCACCATAAGAAGGCTGGAGCTATGAATTCCCTG 1635
Qy 1628 ATACGAGTCTCTG[illegible]GTTCTATCAAATGCTCCTTACCTTCTGAATGTCGATTGTGATCAC 1687
Db 1636 ATACGAGTCTCTG[illegible]GTTCTATCAAATGCTCCTTACCTTCTGAATGTCGATTGTGATCAC 1695
Qy 1688 TACATCAACAATAG[illegible]AAAGCTCTTAGAGAAGC[illegible]AATGTGTTTCATGATGGATCCTCAGTCA 1747
Db 1696 TACATCAACAATAG[illegible]AAAGCTCTTAGAGAAGC[illegible]AATGTGTTTCATGATGGATCCTCAGTCA 1755
Qy 1748 GGAAAGAAAATCTG[illegible]TATGTTCAGTTCCCTCAAAGGTTCGATGGGATTGATAGGCACGAT 1807
Db 1756 GGAAAGAAAATCTG[illegible]TATGTTCAGTTCCCTCAAAGGTTCGATGGGATTGATAGGCACGAT 1815
Qy 1808 CGATACTCAAATCG[illegible]AATGTTGTGTTCTTTGATATCAATATGAAAGGTTTGGATGGGCTA 1867
Db 1816 CGATACTCAAATCG[illegible]AATGTTGTGTTCTTTGATATCAATATGAAAGGTTTGGATGGGCTA 1875
Qy 1868 CAAGGGCCTATATA[illegible]GTCGGTACAGGTTGTGTTTTCAGGAGGCAAGCGCTTTACGGATTT 1927
Db 1876 CAAGGGCCTATATA[illegible]GTCGGTACAGGTTGTGTTTTCAGGAGGCAAGCGCTTTACGGATTT 1935
Qy 1928 GATGCACCGAAGAAGAAGAAGGGCCCACGTAAGACATGCAATTGCTGGCCAAAATGGTGT 1987
Db 1936 GATGCACCGAAGAAGAAGAAGGGCCCACGTAAGACATGCAATTGCTGGCCAAAATGGTGT 1995
Qy 1988 CTCCTATGTTTTGGTTCAAGAAAGAATCGTAAAGCAAAGACAGTGGCTGCGGATAAGAAG 2047
Db 1996 CTCCTATGTTTTGGTTCAAGAAAGAATCGTAAAGCAAAGACAGTGGCTGCGGATAAGAAG 2055
Qy 2048 AAGAAGAATAGGGAAGCGTCAAAGCAGATCCACGCATTAGAAAATATCGAAGAGGGCCGC 2107
Db 2056 AAGAAGAATAGGGAAGCGTCAAAGCAGATCCACGCATTAGAAAATATCGAAGAGGGCCGC 2115
Qy 2108 GGTCATAAAGTTCTTAACGTAGAACAGTCAACCGAGGCAATGCAAATGAAGTTGCAGAAG 2167
Db 2116 GTCACTAAAGGTTCTAACGTAGAACAGTCAACCGAGGCAATGCAAATGAAGTTGGAGAAG 2175
Qy 2168 AAATATGGGCAGTCTCCTGTATTTGTTGCATCTGCGCGTCTGGAGAATGGTGGGATGGCT 2227
Db 2176 AAATTTGGGCAGTCTCCTGTATTTGTTGCATCTGCGCGTATGGAGAATGGTGGGATGGCT 2235
Qy 2228 AGAAACGCAAGCCCGGCTTGTCTGCTTAAAGAAGCCATCCAAGTCATTAGTCGCGGATAT 2287
Db 2236 AGAAACGCAAGCCCGGCTTGTCTGCTTAAAGAAGCCATCCAAGTCATTAGTTGCGGATAT 2295
Qy 2288 GAAGATAAAACTGAATGGGGAAAAGAGATTGGGTGGATCTATGGTTCTGTTACCGAAGAT 2347

```
Db       2296 GAAGATAAAACTGAATGGGGAAAAGAGATTGGGTGGATCTATGGTTCTGTTACCGAAGAT 2355

Qy       2348 ATTCTTACGGGTTCTAAGATGCATTCTCATGGTTGGAGACATGTTTATTGTACACCAAAG 2407

Db       2356 ATTCTTACGGGTTTTAAGATGCATTCTCATGGTTGGAGATCTGTTTATTGTACACCAAAG 2415

Qy       2408 TTAGCGGCTTTCAAAGGATCAGCTCCAATCAATCTTTCGGATCGTCTCCATCAAGTTCTT 2467

Db       2416 TTAGCGGCTTTCAAAGGATCAGCTCCAATCAATCTTTCGGATCGTCTCCATCAAGTTCTT 2475

Qy       2468 CGATGGGCGCTTGGGTCGGTTGAGATTTTCTTGAGTAGGCATTGTCCTATTTGGTATGGT 2527

Db       2476 CGATGGGCGCTTGGGTCGGTTGAGATTTTCTTGAGTAGGCATTGTCCTATTTGGTATGGT 2535

Qy       2528 TATGGAGGTGGGTTGAAATGGCTTGAGCGGTTGTCCTACATTAACTCTGTGGTTTACCCG 2587

Db       2536 TATGGAGGTGGGTTGAAATGGCTTGAGCGGTTGTCCTACATTAACTCTGTGGTTTACCCG 2595

Qy       2588 TGGACCTCTCTACCGCTCATCGTTTACTGTTCTCTCCCTGCCATCTGTCTTCTCACTGGA 2647

Db       2596 TGGACCTCTCTACCGCTCATCGTTTACTGTTCTCTCCCTGCCATCTGTCTTCTCACTGGA 2655

Qy       2648 AAATTCATCGTTCCCGAGATTAGCAACTATGCGAGTATCCTCTTCATGGCGCTCTTCTCG 2707

Db       2656 AAATTCATCGTTCCCGAGATTAGCAACTATGCGAGTATCCTCTTCATGGCGCTCTTCTCG 2715

Qy       2708 TCGATTGCAATAACGGGTATTCTCGAGATGCAATGGGGCAAAGTTGGGATCGATGATTGG 2767

Db       2716 TCGATTGCAATAACGGGTATTCTCGAGATGCAATGGGGCAAAGTTGGGATCGATGATTGG 2775

Qy       2768 TGGAGAAACGAACAGTTTTGGGTCATTGGAGGTGTTTCTGCGCATCTGTTTGCTCTCTTC 2827

Db       2776 TGGAGAAACGAACAGTTTTGGGTCATTGGAGGTGTTTCTGCGCATCTGTTTGCTCTCTTC 2835

Qy       2828 CAAGGTCTCCTCAAGGTTCTTGCTGGTGTCGACACTAACTTCACAGTCACATCAAAAGCA 2887

Db       2836 CAAGGTCTCCTCAAGGTTCTTGCTGGTGTCGACACTAACTTCACAGTCACATCAAAAGCA 2895

Qy       2888 GCTGATGATGGAGAGTTCTCTGACCTTTACCTCTTCAAATGGACTTCACTTCTCATCCCT 2947

Db       2896 GCTGATGATGGAGAGTTCTCTGACCTTTACCTCTTCAAATGGACTTCACTTCTCATCCCT 2955

Qy       2948 CCAATGACTCTACTCATCATAAACGTCATTGGAGTCATAGTCGGAGTCTCTGATGCCATC 3007

Db       2956 CCAATGACTCTACTCATCATAAACGTCATTGGAGTCATAGTCGGAGTCTCTGATGCCATC 3015

Qy       3008 AGCAATGGATACGACTCGTGGGGACCGCTTTTCGGAAGACTGTTCTTTGCACTTTGGGTC 3067

Db       3016 AGCAATGGATACGACTCGTGGGGACCGCTTTTCGGAAGACTGTTCTTTGCACTTTGGGTC 3075

Qy       3068 ATCATTCATCTTTACCCGTTCTTTAAAGACTCCTTCATTGTGTTCATTAGATGATGAAAAA 3127

Db       3076 ATCATTCATCTTTACCCGTTCTTTAAAGACTCCTTCATTGTGTTCATTAGATGATGAAAAA 3135

Qy       3128 ATTATTGTCGTCTGGTCCATCCTCCTGGCCTCGATTCTTACACTTCTTTGGGTCCGGGTT 3187

Db       3136 ATTATTGTCGTCTGGTCCATCCTCCTGGCCTCGATTCTTACACTTCTTTGGGTCCGGGTT 3195

Qy       3188 AATCCGTTTGTGGCGAAAGGCGGTCCTATTCTCGAGATCTGTGGTTTAGACTGCTTGTGA 3247

Db       3196 AATCCGTTTGTGGCGAAAGGCGGTCCTATTCTCGAGATCTGTGGTTTAGACTGCTTGTGA 3255


RESULT 2
US-09-770-444-763/c
; Sequence 763, Application US/09770444
; Patent No. US20020023280A1
; GENERAL INFORMATION:
;  APPLICANT: Gorlach, Jorn
;  APPLICANT:  An, Yong-Qiang
;  APPLICANT:  Hamilton, Carol M.
;  APPLICANT:  Price, Jennifer L.
;  APPLICANT:  Raines, Tracy M.
;  APPLICANT:  Yu, Yang
;  APPLICANT:  Rameaka, Joshua G.
;  APPLICANT:  Page, Amy
;  APPLICANT:  Matthew, Abraham V.
;  APPLICANT:  Ledford, Brooke L.
;  APPLICANT:  Woessner, Jeffrey P.
;  APPLICANT:  Haas, William David
;  APPLICANT:  Garcia, Carlos A.
;  APPLICANT:  Kricker, Maja
;  APPLICANT:  Slader, Ted
;  APPLICANT:  Davis, Keith R.
;  APPLICANT:  Allen, Keith
;  APPLICANT:  Hoffman, Neil
;  APPLICANT:  Hurban, Patrick
;  TITLE OF INVENTION: Expressed Sequences of Arabidopsis
;  TITLE OF INVENTION:  thaliana
;  FILE REFERENCE: 2027 (PARA-016PRV)
;  CURRENT APPLICATION NUMBER: US/09/770,444
;  CURRENT FILING DATE:  2001-01-26
;  PRIOR APPLICATION NUMBER: 60/178,502
;  PRIOR FILING DATE: 2000-01-27
;  NUMBER OF SEQ ID NOS: 999
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 763
;   LENGTH: 447
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
US-09-770-444-763

  Query Match             10.5%;  Score 361;  DB 9;  Length 447;
  Best Local Similarity   100.0%;  Pred. No. 7.9e-185;
  Matches  361;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3076 TCTTTACCCGTTCCTTAAAGGTTTGCTTGGGAAACAAGATAGAATGCCAACCATTATTGT 3135

Db        447 TCTTTACCCGTTCCTTAAAGGTTTGCTTGGGAAACAAGATAGAATGCCAACCATTATTGT 388

Qy       3136 CGTCTGGTCCATCCTCCTGGCCTCGATTCTTACACTTCTTTGGGTCCGGGTTAATCCGTT 3195

Db        387 CGTCTGGTCCATCCTCCTGGCCTCGATTCTTACACTTCTTTGGGTCCGGGTTAATCCGTT 328

Qy       3196 TGTGGCGAAAGGCGGTCCTATTCTCGAGATCTGTGGTTTAGACTGCTTGTGATTCGATTG 3255

Db        327 TGTGGCGAAAGGCGGTCCTATTCTCGAGATCTGTGGTTTAGACTGCTTGTGATTCGATTG 268

Qy       3256 ACCGGTGGATGGGTTGGTGAAAAAGGTTTAATTCCCACGGATCAAAGAGAGGTAAGAGAG 3315

Db        267 ACCGGTGGATGGGTTGGTGAAAAAGGTTTAATTCCCACGGATCAAAGAGAGGTAAGAGAG 208

Qy       3316 ATATTGTTTTACCTCTAAAAGACTCCTTCATTGTGTTCATTAGATGATGAAAAATGAAAA 3375

Db        207 ATATTGTTTTACCTCTAAAAGACTCCTTCATTGTGTTCATTAGATGATGAAAAATGAAAA 148

Qy       3376 GAAAAAGAAGATTTAATTTTGTTACGAGAATTGTTATTTTTGCAAGAATGTGTTGTAGAT 3435

Db        147 GAAAAAGAAGATTTAATTTTGTTACGAGAATTGTTATTTTTGCAAGAATGTGTTGTAGAT 88

Qy       3436 A 3436

Db          87 A 87


RESULT 3
US-09-924-035A-676/c
; Sequence 676, Application US/09924035A
; Patent No. US20020142319A1
; GENERAL INFORMATION:
;  APPLICANT: Grlach, Jrn
;  TITLE OF INVENTION: Expressed Sequences of Arabidopsis
;  TITLE OF INVENTION:  thaliana
;  FILE REFERENCE: 2011US
;  CURRENT APPLICATION NUMBER: US/09/924,035A
;  CURRENT FILING DATE:  2000-08-11
```

PRIOR APPLICATION NUMBER: US 60/148,784
PRIOR FILING DATE: 1999-08-13
NUMBER OF SEQ ID NOS: 900
SOFTWARE: FastSEQ for Windows Version 3.0
SEQ ID NO 676
LENGTH: 342
TYPE: DNA
ORGANISM: Arabidopsis thaliana
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(342)
OTHER INFORMATION: n = A,T,C or G
S-09-924-035A-676

Query Match 6.4%; Score 222; DB 10; Length 342;
Best Local Similarity 100.0%; Pred. No. 2.5e-109;
Matches 222; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
y       3215 ATTCTCGAGATCTGTGGTTTAGACTGCTTGTGATTCGATTGACCGGTGGATGGGTTGGTG 3274
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
b        265 ATTCTCGAGATCTGTGGTTTAGACTGCTTGTGATTCGATTGACCGGTGGATGGGTTGGTG 206

y       3275 AAAAAGGTTTAATTCCCACGGATCAAAGAGAGGTAAGAGAGATATTGTTTTACCTCTAAA 3334
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
b        205 AAAAAGGTTTAATTCCCACGGATCAAAGAGAGGTAAGAGAGATATTGTTTTACCTCTAAA 146

y       3335 AGACTCCTTCATTGTGTTCATTAGATGATGAAAAATGAAAAGAAAAAGAAGATTTAATTT 3394
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
b        145 AGACTCCTTCATTGTGTTCATTAGATGATGAAAAATGAAAAGAAAAAGAAGATTTAATTT 86

y       3395 TGTTACGAGAATTGTTATTTTTGCAAGAATGTGTTGTAGATA 3436
            |||||||||||||||||||||||||||||||||||||||||||
b         85 TGTTACGAGAATTGTTATTTTTGCAAGAATGTGTTGTAGATA 44
```

RESULT 4
S-09-900-237-17
Sequence 17, Application US/09900237
Patent No. US20020120124A1
GENERAL INFORMATION:
APPLICANT: Allen, Stephen
TITLE OF INVENTION: Plant Cellulose Synthases
FILE REFERENCE: BB1170 US CIP
CURRENT APPLICATION NUMBER: US/09/900,237
CURRENT FILING DATE: 2001-07-06
PRIOR APPLICATION NUMBER: 60/092,844
PRIOR FILING DATE: 1998-07-14
PRIOR APPLICATION NUMBER: PCT/US99/15871
PRIOR FILING DATE: 1999-07-13
PRIOR APPLICATION NUMBER: 09/720383
PRIOR FILING DATE: 2000-12-21
NUMBER OF SEQ ID NOS: 33
SOFTWARE: Microsoft Office 97
SEQ ID NO 17
LENGTH: 2890
TYPE: DNA
ORGANISM: Glycine max
S-09-900-237-17

Query Match 0.9%; Score 32; DB 10; Length 2890;
Best Local Similarity 100.0%; Pred. No. 4.9e-06;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1481 GATCATCCTGGCATGATTCAGGTCTTCCTTGG 1512
            ||||||||||||||||||||||||||||||||
Db       614 GATCATCCTGGCATGATTCAGGTCTTCCTTGG 645
```

RESULT 5
US-09-938-842A-559
Sequence 559, Application US/09938842A
Patent No. US20020160378A1
GENERAL INFORMATION:
APPLICANT: Harper, Jeff
APPLICANT: Kreps, Joel
APPLICANT: Wang, Xun
APPLICANT: Zhu, Tong
TITLE OF INVENTION: STRESS REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
TITLE OF INVENTION: SAME, AND METHODS OF USE
FILE REFERENCE: SCRIP1300.3
CURRENT APPLICATION NUMBER: US/09/938,842A
CURRENT FILING DATE: 2001-08-24
PRIOR APPLICATION NUMBER: US 60/227,866
PRIOR FILING DATE: 2000-08-24
PRIOR APPLICATION NUMBER: US 60/264,647
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: US 60/290,111
PRIOR FILING DATE: 2001-05-22
NUMBER OF SEQ ID NOS: 691
SEQ ID NO 559
LENGTH: 3255
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-938-842A-559

Query Match 0.9%; Score 32; DB 10; Length 3255;
Best Local Similarity 100.0%; Pred. No. 4.9e-06;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       419 TTGGATTCAGCTCCACCTGGCTCTCAGATTCC 450
             ||||||||||||||||||||||||||||||||
Db       415 TTGGATTCAGCTCCACCTGGCTCTCAGATTCC 446
```

RESULT 6
US-10-229-193-7
Sequence 7, Application US/10229193
Publication No. US20030126643A1
GENERAL INFORMATION:
APPLICANT: Arioli, Antonio Williamson, Richard E.
APPLICANT: Betzner, Andreas S.
APPLICANT: Peng, Liangcai
TITLE OF INVENTION: Manipulation of Cellulose and/or Beta-1,4-Glucan
FILE REFERENCE: 96-98A
CURRENT APPLICATION NUMBER: US/10/229,193
CURRENT FILING DATE: 2002-08-26
PRIOR APPLICATION NUMBER: 09/221,013
PRIOR FILING DATE: 1998-12-23
PRIOR APPLICATION NUMBER: PCT/AU97/00402
PRIOR FILING DATE: 1997-06-24
PRIOR APPLICATION NUMBER: AU PO0699
PRIOR FILING DATE: 1996-06-27
NUMBER OF SEQ ID NOS: 36
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 7
LENGTH: 3828
TYPE: DNA
ORGANISM: Arabidopsis thaliana
FEATURE:
NAME/KEY: CDS
LOCATION: (239)..(3490)
US-10-229-193-7

Query Match 0.9%; Score 32; DB 14; Length 3828;
Best Local Similarity 100.0%; Pred. No. 5e-06;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       419 TTGGATTCAGCTCCACCTGGCTCTCAGATTCC 450
             ||||||||||||||||||||||||||||||||
Db       653 TTGGATTCAGCTCCACCTGGCTCTCAGATTCC 684
```

RESULT 7
US-10-229-193-4
Sequence 4, Application US/10229193
Publication No. US20030126643A1

; GENERAL INFORMATION:
; APPLICANT: Arioli, Antonio Williamson, Richard E.
; APPLICANT: Betzner, Andreas S.
; APPLICANT: Peng, Liangcai
; TITLE OF INVENTION: Manipulation of Cellulose and/or Beta-1,4-Glucan
; FILE REFERENCE: 96-98A
; CURRENT APPLICATION NUMBER: US/10/229,193
; CURRENT FILING DATE: 2002-08-26
; PRIOR APPLICATION NUMBER: 09/221,913
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: PCT/AU97/00402
; PRIOR FILING DATE: 1997-06-24
; PRIOR APPLICATION NUMBER: AU PO0693
; PRIOR FILING DATE: 1996-06-27
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 5009
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(5009)
; OTHER INFORMATION: N is A, T, C or G.
US-10-229-193-4

Query Match 0.9%; Score 32; DB 14; Length 5009;
Best Local Similarity 100.0%; Pred. No. 5.1e-06;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 419 TTGGATTCAGCTCCACCTGGCTCTCAGATTCC 450
||||||||||||||||||||||||||||||||
Db 1787 TTGGATTCAGCTCCACCTGGCTCTCAGATTCC 1818

RESULT 8
US-09-770-961-707/c
; Sequence 707, Application US/09770961
; Publication No. US20030115639A1
; GENERAL INFORMATION:
; APPLICANT: Gorlach, Jorn
; APPLICANT: An, Yong-Qiang
; APPLICANT: Hamilton, Carol M.
; APPLICANT: Price, Jennifer L.
; APPLICANT: Raines, Tracy M.
; APPLICANT: Yu, Yang
; APPLICANT: Ramaka, Joshua G.
; APPLICANT: Page, Amy
; APPLICANT: Matthew, Abraham V.
; APPLICANT: Ledford, Brooke L.
; APPLICANT: Woessner, Jeffrey P.
; APPLICANT: Haas, William David
; APPLICANT: Garcia, Carlos A.
; APPLICANT: Kricker, Maja
; APPLICANT: Slader, Ted
; APPLICANT: Davis, Keith R.
; APPLICANT: Allen, Keith
; APPLICANT: Hoffman, Neil
; APPLICANT: Hurban, Patrick
; TITLE OF INVENTION: Expressed Sequences of Arabidopsis
; TITLE OF INVENTION: thaliana
; FILE REFERENCE: 2026 (PARA-015PRV)
; CURRENT APPLICATION NUMBER: US/09/770,961
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: 60/178,466
; PRIOR FILING DATE: 2000-01-27
; NUMBER OF SEQ ID NOS: 999
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 707
; LENGTH: 489
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-770-961-707

Query Match 0.8%; Score 29; DB 11; Length 489;
Best Local Similarity 100.0%; Pred. No. 0.00018;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3188 AATCCGTTTGTGGCGAAAGGCGGTCCTAT 3216
|||||||||||||||||||||||||||||
Db 259 AATCCGTTTGTGGCGAAAGGCGGTCCTAT 231

RESULT 9
US-10-209-059-27
; Sequence 27, Application US/10209059
; Publication No. US20030163838A1
; GENERAL INFORMATION:
; APPLICANT: Dhugga, Kanwarpal S.
; APPLICANT: Wang, Haiyin
; TITLE OF INVENTION: Maize Cellulose Synthases and Uses
; TITLE OF INVENTION: Thereof
; FILE REFERENCE: 0864R2
; CURRENT APPLICATION NUMBER: US/10/209,059
; CURRENT FILING DATE: 2002-07-31
; PRIOR APPLICATION NUMBER: 60/096,822
; PRIOR FILING DATE: 1998-08-17
; PRIOR APPLICATION NUMBER: 09/371,383
; PRIOR FILING DATE: 1999-08-06
; PRIOR APPLICATION NUMBER: 09/550,483
; PRIOR FILING DATE: 2000-04-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 3231
; TYPE: DNA
; ORGANISM: Zea mays
US-10-209-059-27

Query Match 0.8%; Score 29; DB 12; Length 3231;
Best Local Similarity 100.0%; Pred. No. 0.00021;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1070 AGTACAGTGGATCCATTGAAAGAGCCTCC 1098
|||||||||||||||||||||||||||||
Db 816 AGTACAGTGGATCCATTGAAAGAGCCTCC 844

RESULT 10
US-09-838-539-1
; Sequence 1, Application US/09838539
; Patent No. US20020129401A1
; GENERAL INFORMATION:
; APPLICANT: Stalker, D. et al.
; TITLE OF INVENTION: Plant Cellulose Synthase and Promoter
; TITLE OF INVENTION: Sequences
; FILE REFERENCE: 15621/03/US
; CURRENT APPLICATION NUMBER: US/09/838,539
; CURRENT FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/029,987
; PRIOR FILING DATE: 1996-10-29
; PRIOR APPLICATION NUMBER: 08/960,048
; PRIOR FILING DATE: 1997-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 3328
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Oligonucleotide
US-09-838-539-1

Query Match 0.8%; Score 29; DB 10; Length 3328;
Best Local Similarity 100.0%; Pred. No. 0.00021;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
y       1070 AGTACAGTGGATCCATTGAAAGAGCCTCC 1098
             ||||||||||||||||||||||||||||
b        947 AGTACAGTGGATCCATTGAAAGAGCCTCC 975


ESULT 11
S-10-209-059-13
 Sequence 13, Application US/10209059
 Publication No. US20030163838A1
 GENERAL INFORMATION:
  APPLICANT: Dhugga, Kanwarpal S.
  APPLICANT:  Wang, Haiyin
  TITLE OF INVENTION: Maize Cellulose Synthases and Uses
  TITLE OF INVENTION:  Thereof
  FILE REFERENCE: 0864R2
  CURRENT APPLICATION NUMBER: US/10/209,059
  CURRENT FILING DATE:  2002-07-31
  PRIOR APPLICATION NUMBER: 60/096,822
  PRIOR FILING DATE: 1998-08-17
  PRIOR APPLICATION NUMBER: 09/371,383
  PRIOR FILING DATE: 1999-08-06
  PRIOR APPLICATION NUMBER: 09/550,483
  PRIOR FILING DATE: 2000-04-14
  NUMBER OF SEQ ID NOS: 52
  SOFTWARE: FastSEQ for Windows Version 4.0
 SEQ ID NO 13
   LENGTH: 3704
   TYPE: DNA
   ORGANISM: Zea mays
S-10-209-059-13

 Query Match             0.8%;  Score 29;  DB 12;  Length 3704;
 Best Local Similarity   100.0%;  Pred. No. 0.00021;
 Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

y       2312 GAGATTGGGTGGATCTATGGTTCTGTTAC 2340
             ||||||||||||||||||||||||||||
b       2564 GAGATTGGGTGGATCTATGGTTCTGTTAC 2592


ESULT 12
S-10-160-719-57
 Sequence 57, Application US/10160719
 Publication No. US20030167528A1
 GENERAL INFORMATION:
  APPLICANT: Kanwarpal S. Dhugga
  APPLICANT:  Timothy G. Helentjaris
  APPLICANT:  Benjamin A.  Bowen
  APPLICANT:  Xun Wang
  TITLE OF INVENTION: Maize Cellulose Synthases and Uses
  TITLE OF INVENTION:  Thereof
  FILE REFERENCE: 0864
  CURRENT APPLICATION NUMBER: US/10/160,719
  CURRENT FILING DATE:  2002-06-03
  PRIOR APPLICATION NUMBER: US/09/371,383A
  PRIOR FILING DATE: 1999-08-06
  PRIOR APPLICATION NUMBER: US 60/096,822
  PRIOR FILING DATE: 1998-08-17
  NUMBER OF SEQ ID NOS: 60
  SOFTWARE: FastSEQ for Windows Version 3.0
 SEQ ID NO 57
   LENGTH: 3704
   TYPE: DNA
   ORGANISM: Zea mays
   FEATURE:
   NAME/KEY: CDS
   LOCATION: (272)...(3499)
S-10-160-719-57

 Query Match             0.8%;  Score 29;  DB 12;  Length 3704;
 Best Local Similarity   100.0%;  Pred. No. 0.00021;
 Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2312 GAGATTGGGTGGAT TATGGTTCTGTTAC 2340
             || ||| ||| . .||| ||| ||||
Db      2564 GAGATTGGGTGGAT TATGGTTCTGTTAT 2592


RESULT 13
US-09-900-237-13
; Sequence 13, Application [illegible] 09900237
; Patent No. US20020120124A1
; GENERAL INFORMATION:
;  APPLICANT: Allen, Stephen
;  TITLE OF INVENTION: Plant Cellulose Synthases
;  FILE REFERENCE: BB1170 US CIP
;  CURRENT APPLICATION NUMBER: US/09/900,237
;  CURRENT FILING DATE: 2001-07-06
;  PRIOR APPLICATION NUMBER: 60/217,846
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: [illegible]
;  PRIOR FILING DATE: 1999-12-13
;  PRIOR APPLICATION NUMBER: 09/520,464
;  PRIOR FILING DATE: 2000-03-08
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 13
;   LENGTH: 3517
;   TYPE: DNA
;   ORGANISM: Glycine max
US-09-900-237-13

  Query Match             0.8%;  Score 27;  DB 10;  Length 3517;
  Best Local Similarity   100.0%;  Pred. No. 0.0026;
  Matches   27;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2282 GGATATGAAGATAAAACTGAATGGGGA 2308
             |||||| ||| ||||||||||||||||
Db      2310 GGATATGAAGATAAAACTGAATGGGGA 2336


RESULT 14
US-10-209-059-45
; Sequence 45, Application US/10209059
; Publication No. US20030163838A1
; GENERAL INFORMATION:
;  APPLICANT: Dhugga, Kanwarpal S.
;  APPLICANT:  Wang, Haiyin
;  TITLE OF INVENTION: Maize Cellulose Synthases and Uses
;  TITLE OF INVENTION:  Thereof
;  FILE REFERENCE: 0864R2
;  CURRENT APPLICATION NUMBER: US/10/209,059
;  CURRENT FILING DATE:  2002-07-31
;  PRIOR APPLICATION NUMBER: 60/096,822
;  PRIOR FILING DATE: 1998-08-17
;  PRIOR APPLICATION NUMBER: 09/371,383
;  PRIOR FILING DATE: 1999-08-06
;  PRIOR APPLICATION NUMBER: 09/550,483
;  PRIOR FILING DATE: 2000-04-14
;  NUMBER OF SEQ ID NOS: 52
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 45
;   LENGTH: 3725
;   TYPE: DNA
;   ORGANISM: Zea mays
US-10-209-059-45

  Query Match             0.8%;  Score 26;  DB 12;  Length 3725;
  Best Local Similarity   100.0%;  Pred. No. 0.009;
  Matches   26;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1700 AGCAAAGCTCTTAGAGAAGCAATGTG 1725
             ||||||||||||||||||||||||||
Db      1868 AGCAAAGCTCTTAGAGAAGCAATGTG 1893
```

```
RESULT 15
US-10-160-719-13
; Sequence 13, Application US/10160719
; Publication No. US20030167528A1
; GENERAL INFORMATION:
;  APPLICANT: Kanwarpal S. Dhugga
;  APPLICANT:  Timothy G. Helentjaris
;  APPLICANT:  Benjamin A.  Bowen
;  APPLICANT:  Xun Wang
;  TITLE OF INVENTION: Maize Cellulose Synthases and Uses
;  TITLE OF INVENTION:  Thereof
;  FILE REFERENCE: 0864
;  CURRENT APPLICATION NUMBER: US/10/160,719
;  CURRENT FILING DATE:  2002-06-03
;  PRIOR APPLICATION NUMBER: US/09/371,383A
;  PRIOR FILING DATE: 1999-08-06
;  PRIOR APPLICATION NUMBER: US 60/096,822
;  PRIOR FILING DATE: 1998-08-17
;  NUMBER OF SEQ ID NOS: 60
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 13
;   LENGTH: 3725
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (179)...(3400)
US-10-160-719-13

  Query Match             0.8%;  Score 26;  DB 12;  Length 3725;
  Best Local Similarity   100.0%;  Pred. No. 0.009;
  Matches   26;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1700 AGCAAAGCTCTTAGAGAAGCAATGTG 1725
             ||||||||||||||||||||||||||
Db       1868 AGCAAAGCTCTTAGAGAAGCAATGTG 1893


Search completed: October 31, 2003, 15:45:51
Job time : 859.389 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 29, 2003, 18:26:33 ; Search time 31.2128 Seconds
(without alignments)
3013.285 Million cell updates/sec

Title: US-09-980-043-2
Perfect score: 5240
Sequence: 1 MMESGAPICHTCGEQVGHDA..... .....PFVNKVLSILVAEYCISIDC 978

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_76:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4524	86.3	974	2	T10797	cellulose synthase
2	4202	80.2	958	2	T04870	cellulose synthase
3	3347	63.9	1081	2	T05351	cellulose synthase
4	3346	63.9	1026	2	T51579	cellulose synthase
5	3327	63.5	1069	2	T52084	cellulose synthase
6	3274	62.5	1084	2	T08584	cellulose synthase
7	3251	62.0	1065	2	F84669	probable cellulose
8	3248	62.0	1088	2	H84614	probable cellulose
9	3199.5	61.1	1081	2	T52018	cellulose synthase
10	2730	52.1	685	2	T10800	cellulose synthase
11	2233.5	42.6	1145	2	T51546	cellulose synthase
12	2157.5	41.2	1111	2	T05646	hypothetical prote
13	2142	40.9	1181	2	D86187	hypothetical prote
14	2080.5	39.7	1036	2	D84741	probable cellulose
15	2059.5	39.3	979	2	C86446	probable cellulose
16	1565.5	29.9	583	2	T02209	cellulose synthase
17	1057.5	20.2	712	2	T02562	cellulose synthase
18	1048.5	20.0	757	2	T02561	probable cellulose
19	1035.5	19.8	755	2	T02563	cellulose synthase
20	989	18.9	748	2	T02560	cellulose synthase
21	956.5	18.3	686	2	T08919	hypothetical prote
22	937.5	17.9	689	2	T08918	hypothetical prote
23	929.5	17.7	727	2	T08920	hypothetical prote
24	909	17.3	828	2	E71417	hypothetical prote
25	837	16.0	710	2	B71417	hypothetical prote
26	416.5	7.9	693	2	AF2275	cellulose synthase
27	267	5.1	759	2	D70422	cellulose synthase
28	230	4.4	874	2	AB0985	probable polysacch
29	223	4.3	754	2	A43735	bcsA protein - Ace
30	222	4.2	888	2	E91180	probable cellulose
31	221	4.2	692	2	S47754	hypothetical prote
32	221	4.2	888	2	H65151	hypothetical 101.6
33	215	4.1	888	2	G86026	probable cellulose
34	207.5	4.0	861	2	I39714	cellulose synthase
35	207.5	4.0	1596	2	T31338	cellulose synthase
36	205	3.9	729	2	A98320	cellulose synthase
37	205	3.9	729	2	AD2963	cellulose synthase
38	188.5	3.6	326	2	T08591	TGACG-motif bindin
39	180.5	3.4	654	2	AB2932	beta 1,3 glucan sy
40	180.5	3.4	654	2	D98350	hypothetical prote
41	176	3.4	322	2	T12093	TGACG-motif bindin
42	174.5	3.3	723	2	C36963	cellulose synthase
43	172.5	3.3	322	2	T08592	TGACG-motif-bindin
44	160	3.1	768	2	B97083	glycosyltransferas
45	155	3.0	664	2	A95889	probable cellulose

ALIGNMENTS

RESULT 1
T10797
cellulose synthase (EC 2.4.1.-) catalytic chain celA1 - upland cotton
C;Species: Gossypium hirsutum (upland cotton)
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
C;Accession: T10797
R;Pear, J.R.; Kawagoe, Y.; Schreckengost, W.E.; Delmer, D.P.; Stalker, D.M.
Proc. Natl. Acad. Sci. U.S.A. 93, 12637-12642, 1996
A;Title: Higher plants contain homologs of the bacterial celA genes encoding the cata
A;Reference number: Z17152; MUID:97057296; PMID:8901635
A;Accession: T10797
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-974 <PEA>
A;Cross-references: EMBL:U58283; NID:g1706955; PIDN:AAB37766.1; PID:g1706956
A;Experimental source: strain Acala SJ-2; fiber
C;Genetics:
A;Gene: celA1
C;Function:
A;Description: involved in the synthesis of cellulose
C;Keywords: glycosyltransferase; hexosyltransferase

Query Match 86.3%; Score 4524; DB 2; Length 974;
Best Local Similarity 85.5%; Pred. No. 0;
Matches 840; Conservative 70; Mismatches 60; Indels 12; Gaps 7;

```
Qy          1 MMESGAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEINEGRKVCLRCGSPYDE 60
              ||||| |:|||||| || : ||| |||||||::|:|||||||:: |||| ||||||||||
Db          1 MMESGVPVCHTCGEHVGLNVNGEPFVACHECNFPICKSCFEYDLKEGRKACLRCGSPYDE 60

Qy         61 NLLDDVEKKGSGNQSTMASHLNDSQDVGIHARHISSVSTVDSEMNDEYGNPIWKNRVKSC 120
              |:|||||  | :|:|||||:||| |||||||||||||||:|||| :: || ||||||:|
Db         61 NLLDDVE-KATGDQSTMAAHLNKSQDVGIHARHISSVSTLDSEMAEDNGNSIWKNRVESW 119

Qy        121 KDKENKKKKRSPKAETEPAQVPTEQQMEEKPSAEASEPLSIVYPIPRNKLTPYRAVIIMR 180
              |:|:||||| :       |::| |||||:||: :||:||| : |||:::| ||| |||||
Db        120 KEKKNKKKKPATTKVEREAEIPPEQQMEDKPAPDASQPLSTIIPIPKSRLAPYRTVIIMR 179

Qy        181 LVILGLFFHFRITNPVDSAFGLWLTSVICEIWFAFSWVLDQFPKWNPVNRETYIERLSAR 240
              |:|||||:|:|||||||||||||||||||||||||||||||| |||||||:|||||
Db        180 LIILGLFFHYRVTNPVDSAFGLWLTSVICEIWFAFSWVLDQFPKWYPVNRETYIDRLSAR 239

Qy        241 YEREGEPSQLAGVDFFVSTVDPLKEPPLITANTVLSILAVDYPVDKVSCYVSDDGAAMLS 300
              ||||||| :|| ||||||||||||||||||||||||||:||||||||||:||||||||:
Db        240 YEREGEPDELAAVDFFVSTVDPLKEPPLITANTVLSILALDYPVDKVSCYISDDGAAMLT 299

Qy        301 FESLVETAEFARKWVPFCKKFSIEPRAPEFYFSQKIDYLKDKVQPSFVKERRAMKRDYEE 360
              ||||||||:|||||||||||||||||||||||||||||||||||||||||||||||||||
Db        300 FESLVETADFARKWVPFCKKFSIEPRAPEFYFSQKIDYLKDKVQPSFVKERRAMKRDYEE 359

Qy        361 YKVRVNALVAKAQKTPEEGWTMQDGTPWPGNNTRDHPGHDSGLPWEIL---GARDIEGNE 417
              ||:|:|||||||||||:'|||||||| |||||  ||||     |:     |   |||||||||
```

```
o        360 YKIRINALVAKAQKTPDEGWTMQDGTSWPGNNPRDHP----GMIQVFLGYSGARDIEGNE 415

y        418 LPRLVYVSREKRPGYQHHKKAGAENALVRVSAVLTNAPYILNVDCDHYVNNSKAVREAMC 477

o        416 LPRLVYVSREKRPGYQHHKKAGAENALVRVSAVLTNAPFILNLDCDHYVNNSKAVREAMC 475

         478 ILMDPQVGRDVCYVQFPQRFDGIDKSDRYANRNVVFFDVNMKGLDGIQGPVYVGTGCVFN 537

o        476 FLMDPQVGRDVCYVQFPQRFDGIDRSDRYANRNTVFFDVNMKGLDGIQGPVYVGTGCVFN 535

y        538 RQALYGYGPPSMPSLRKRKDSSSCFSCCCPSKKKPAQDPAEVYRDAKREDLNAAIFNLTE 597

o        536 RQALYGYGPPSMPSFPK-SSSSSC-SCCCPGKKEP-KDPSELYRDAKREELDAAIFNLRE 592

y        598 IDNYDEHERSMLISQLSFEKTFGLSSVFIESTLMENGGVPESANSPPFIKEAIQVIGCGY 657

o        593 IDNYDEYERSMLISQTSFEKTFGLSSVFIESTLMENGGVAESANPSTLIKEAIHVISCGY 652

y        658 EEKTEWGKQIGWIYGSVTEDILSGFKMHCRGWRSIYCMPVRPAFKGSAPINLSDRLHQVL 717

c        653 EEKTAWGKEIGWIYGSVTEDILTGFKMHCRGWRSIYCMPLRPAFKGSAPINLSDRLHQVL 712

y        718 RWALGSVEIFFSRHCPLWYGFGGGRLKWLQRLAYINTIVYPFTSLPLIAYCTIPAVCLLT 777

o        713 RWALGSVEIFLSRHCPLWYGFGGGRLKWLQRLAYINTIVYPFTSLPLIAYCSLPAICLLT 772

y        778 GKFIIPTLSNLASMLFLGLFISIIVTAVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAV 837

o        773 GKFIIPTLSNLASVLFLGLFLSIIVTAVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAV 832

y        838 FQGFLKMLAGIDTNFTVTAKAAEDAEFGELYMVKWTTLLIPPTTLLIINMSG-CAGFSDA 896

o        833 FQGFLKMLAGIDTNFTVTAKAADDADFGELYIVKWTTLLIPPTTLLIVNMVGVVAGFSDA 892

y        897 LNKGYEAWGPLFGKVFFAFWVILHLYPFLKGLMGRQNLTPTIVVLWSVLLASVFSLVWVK 956

o        893 LNKGYEAWGPLFGKVFFSFWVILHLYPFLKGLMGRQNRTPTIVVLWSVLLASVFSLVWVR 952

y        957 INPFVNKVDNTLVAETCISIDC 978

b        953 INPFVSTADSTTVSQSCISIDC 974


ESULT 2
04870
ellulose synthase (EC 2.4.1.-) catalytic chain F28A21.190 - Arabidopsis thaliana
,Species: Arabidopsis thaliana (mouse-ear cress)
,Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-Jun-1999
,Accession: T04870
;Bevan, M.; Mueller, M.W.; Muendlein, A.; Felber, R.; Bancroft, I.; Mewes, H.W.; Mayer,
ubmitted to the Protein Sequence Database, February 1999
,Reference number: Z15387
;Accession: T04870
;Molecule type: DNA
;Residues: 1-958 <BEV>
;Cross-references: EMBL:AL035526
;Experimental source: cultivar Columbia; BAC clone F28A21
,Genetics:
;Map position: 4
;Introns: 60/1; 76/3; 94/1; 149/2; 240/2; 355/3; 401/3; 443/3; 514/3; 662/3; 780/3; 791
;Note: F28A21.190
;Keywords: glycosyltransferase; hexosyltransferase

 Query Match             80.2%;  Score 4202;  DB 2;  Length 958;
 Best Local Similarity   79.1%;  Pred. No. 1.8e-317;
 Matches  784;  Conservative   85;  Mismatches   76;  Indels   46;  Gaps    9;

y          1 MMESGAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEINEGRKVCLRCGSPYDE 60

b          1 MMESRSPICNTCGEEIGVKSNGEFFVACHECSFPICKACLEYEFKEGRRICLRCGNPYDE 60

y         61 NLLDDVEKKGSGNQSTMASHLNDSQDVGIHARHISSVSTVDSEMNDEYGNPIWKNRVKSC 120

Db         61 NVFDDVETETSKTQ--T          DSGIHAFHISTVSTIDSELNDEYGNPIWKNRVESW 119

Qy        121 ---KDKFNKKKKRS-    KAETFFAQVFTEQQMEEKP---SAEASEPLSIVYPIPRNKLT 171

Db        120 KDKKDKKSKKKKKPVKATKAEQHFAQIFIQQHMEDTPPNTESGATDVLSVVIPIPRTKIT 179

Qy        172 PYRAVIIMRLVILGLFFHFRITNPVHDAFGLWLTSVICEIWFAFSWVLDQFPKWNPVNRE 231

Db        172 SYRIVIIMRLIILALFFNYRITHPVHDAYGLWLTSVICEIWFAVSWVLDQFPKWSPINRE 231

Qy        232 TYIERLSARYEREGEPSQLAAVDFFVSTVDPLKEPPLITANTVLSILAVDYPVDKVSCYV 291

Db        232 TYIDRLSARFEREGEPSQLAAVDFFVSTVDPLKEPPLITANTVLSILALDYPVDKVSCYV 291

Qy        292 SDDGAAMLSFESLVETAEFARKWVPFCKKFSIEPRAPEFYFSQKIDYLKDKVQPSFVKER 351

Db        292 SDDGAAMLSFESIVETADFARKWVPFCKKYSIEPRAPEFYFSLKIDYLRDKVQPSFVKER 351

Qy        352 RAMKREYEEFKVRINALVAKAQKTPDEGWTMQDGTSWPGNNTRDHPGMIQVFLGYS 407

Db        352 RAMKREYEEFKIRINALVAKAQKTPDEGWTMQDGTSWPGNNTRDHP-GMIQVFLGYS 407

Qy        408 GARDIEGNELPRLVYVSREKRPGYQHHKKAGAENALVRVSAVLTNAPYILNVDCDHYVNN 467

Db        408 GARDIEGNELPRLVYVSREKRPGYQHHKKAGAENALVRVSAVLTNAPFILNLDCDHYVNN 467

Qy        468 SKAVREAMCILMDPQVGRDVCYVQFPQRFDGIDKSDRYANRNVVFFDVNMKGLDGIQGPV 527

Db        468 SKAVREAMCFLMDPQVGQDVCFVQFPQRFDGIDKSDRYANRNIVFFDVNMRGLDGIQGPV 527

Qy        528 YVGTGCVFNRQALYGYGPPSMPSLRKRKDSSSCFSCCCPSKKKPAQDPAEVYRDAKREDL 587

Db        528 YVGTGTVFRRQALYGYSPPSKPRILPQSSSS   SCCCLTKKKQPQDPSEIYKDAKREEL 584

Qy        588 NAAIFNLTEIDNYDEHERSMLISQLSFEKTFGLSSVFIESTLMENGGVPESANSPPFIKE 647

Db        585 DAAIFNLGDLDNYDEYDRSMLISQTSFEKTFGLSTVFIESTLMENGGVPDSVNPSTLIKE 644

Qy        648 AIQVIGCGYEEKTEWGKQIGWIYGSVTEDILSGFKMHCRGWRSIYCMPVRPAFKGSAPIN 707

Db        645 AIHVISCGYEEKTEWGKEIGWIYGSITEDILTGFKMHCRGWRSIYCMPLRPAFKGSAPIN 704

Qy        708 LSDRLHQVLRWALGSVEIFFSRHCPLWYGFGGGRLKWLQRLAYINTIVYPFTSLPLIAYC 767

Db        705 LSDRLHQVLRWALGSVEIFLSRHCPLWYGCSGGRLKLLQRLAYINTIVYPFTSLPLVAYC 764

Qy        768 TIPAVCLLTGKFIIPTLSNLASMLFLGLFISIVTAVLELRWSGVSIEDLWRNEQFWVIG 827

Db        765 TLPAICLLTGKFIIPTLSNLASMLFL ---------- GVSIEDLWRNEQFWVIG 807

Qy        828 GVSAHLFAVFQGFLKMLAGIDTNFTVTAKAAEDAEFGELYMVKWTTLLIPPTTLLIINMS 887

Db        808 GVSAHLFAVFQGFLKMLAGLDTNFTVTSKTADDLEFGELYIVKWTTLLIPPTSLLIINLV 867

Qy        888 G-CAGFSDALNKGYEAWGPLFGKVFFAFWVILHLYPFLKGLMGRQNLTPTIVVLWSVLLA 946

Db        868 GVVAGFSDALNKGYEAWGPLFGKVFFAFWVILHLYPFLKGLMGRQNRTPTIVILWSILLA 927

Qy        947 SVFSLVWVKINPFVNKVDNTLVAETCISIDC 978

Db        928 SVFSLVWVRINPFVSKTDTTSLSLNQLLIDC 958


RESULT 3
T05351
cellulose synthase (EC 2.4.1.-) catalytic chain RSW1 - Arabidopsis thaliana
N;Alternate names: protein F9B4.110
C;Species: Arabidopsis thaliana (mouse ear cress)
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-Jun-1999
C;Accession: T05351
R;Bevan, M.; Terryn, N.; Ardiles, W.; Buysshaert, C.; Dasseville, R.; De Clerck, R.; De
ewes, H.W.; Mayer, K.F.X.; Schueller, C.
submitted to the Protein Sequence Database, February 1999
```

A;Reference number: Z15409
A;Accession: T05351
A;Molecule type: DNA
A;Residues: 1-1081 <BEV>
A;Cross references: EMBL:AL034567
A;Experimental source: cultivar Columbia; BAC clone F8B4
C;Genetics:
A;Gene: RSW1
A;Map position: 4
A;Introns: 27/3; 93/1; 150/3; 189/1; 252/2; 341/2; 456/1; 502/3; 544/3; 615/3; 704/1; 76
A;Note: F8B4.110
C;Keywords: glycosyltransferase; hexosyltransferase; P-loop

Query Match 63.9%; Score 3347; DB 2; Length 1081;
Best Local Similarity 61.3%; Pred. No. 4.8e-251;
Matches 644; Conservative 122; Mismatches 176; Indels 108; Gaps 18;

```
Qy          8 ICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEINEGRKVCLRC--------GSP-- 57
              ||  ||: ||    |:::|||:|||::|:|: |:|:|  :| : | :|        |||
Db         38 ICQICGDDVGLAETGDVFVACNECAFPVCRPCYEYERKDGTQCCPQCKTRFRRHRGSPRV 97

Qy         58 - -YDENLLDDVEKKGSGNQ-STMASHLNDSQDVGIHARHIS----------SVST 99
                ||: :|||:| : :  : :  : |  :     ::    :||           : |
Db         98 EGDEDEDDVDDIENEFNYAQGANKAPHQRHGEEFSSSSRHEDQPIPLLTHGHTVSGEIRT 157

Qy        100 VDSE----------------MNDEY----- --------- -----  GNPIWKNRVKSC 120
                |::              ::  |                        ||  || |::
Db        158 PDTQSVRTTSGPLGPSDRNAISSPYIDPRQPVPVRIVDPSFMLNSYGLGNVDWKERVEGW 217

Qy        121 KDKENK-------KKKRSPKAETEPAQVPTEQQMEEKPSAEASE-PLSIVYPIPRNKLTP 172
              | ': |       |       | |     |    ||    |: :  |:! | ||| ::|||
Db        218 KLKQEKNMLQMTGKYHEGKGGEIE----GTGSNGEELQMAIDTRLPMSRVVPIPSSRLTP 273

Qy        173 YRAVIIMRLVILGLFFHFRITNPVDSAFGLWLTSVICEIWFAFSWVLDQFPKWNPVNRET 232
              || |||:|:||  | :| :|:||  :|. |||||||||:|||:|:|:|||| |:|||
Db        274 YRVVIILRLIILCFFLQYRTTHPVKNAYPLWLTSVICEIWFAFSWLLDQFPKWYPINRET 333

Qy        233 YIERLSARYEREGEPSQLAGVDFFVSTVDPLKEPPLITANTVLSILAVDYPVDKVSCYVS 292
              |::||: ||:|:|||||  :| |||:|||||:||||:|||:|||:|:|||||:||||
Db        334 YLDRLAIRYDRDGEPSQLVPVDVFVSTVDPLKEPPLVTANTVLSILSVDYPVDKVACYVS 393

Qy        293 DDGAAMLSFESLVETAEFARKWVPFCKKFSIEPRAPEFYFSQKIDYLKDKVQPSFVKERR 352
              |||:||:|||| :|||||:|||||:|:|| |||:||  |||||||||:|||||||||||
Db        394 DDGSAMLTFESLSETAEFAKKWVPFCKKFNIEPRAPEFYFAQKIDYLKDKIQPSFVKERR 453

Qy        353 AMKRDYEEYKVRVNALVAKAQKTPEEGWTMQDGTPWPGNNTRDHP-------GHDSGLPW 405
              |||:|||:|||: ||||||:' ||||:|||:|||||| |||:.|        ||  ||
Db        454 AMKREYEEFKVRINALVAKAQKIPEEGWTMQDGTPWPGNNTRDHPGMIQVFLGHSGGL-- 511

Qy        406 EILGARDIEGNELPRLVYVSREK-PQFQHFKRAGALNAMVKVPAVLTNAPYILNVDCDHY 465
                  | :||.|.||:||   |.|:  :|:.  : |:  :      |:|  |:|| |||:
Db        512       -DTQGNELPRLIYVSREKRPQFQHQKKAGALNAMVRVSAVLTNAPYLLNVDCDHY 565

Qy        466 VNNSKAVREAMCILMDPQVGRDVCYVQFPQRFDGIDKSDRYANRNVVFFDVNMKGLDGIQ 525
              ||||:::|||| :|| :|:  |||:: ||||| |    :.| |:|||:|||||||| |
Db        566 FNNSKAIKEAMCFMMDPAIGKKCCYVQFPQRFDGIDLHDRYANRNIVFFDINMKGLDGIQ 625

Qy        526 GPVYVGTGCVFNRQALYGYGPPSMPSLRKPKDSSSCFSCC-SKKKPAQDPAEVY - RD 582
              |||||||| ||||:|||| |     : :  :      :| |:||      |   |
Db        626 GPVYVGTGCCFNRQALYGYDP---VLTEELLEPNIIVKSC-GSRKKGKSSKKYNYEKRRG 682

Qy        583 AKREDLNAAIFNLTEID- - NYTPHERSMLISQLSFEKTF GLSSVFIESTLMENGGVPE 638
              | | :| :||: :||     ||: |||:.:|| : ||    |  | :' ||   :|
Db        683 INRSDSNAPLFNMEDIDEGFEGYDD-ERSILMSQRSVEKRF QSPVFIAATFMEQGGIPP 741

Qy        639 SANSPPFIKEAIQVIGCGYEEKTEWGKQIGWIYGSVTEDIL GFKMHCRGWRSIYCMPVR 698
              : |    :||| || ||||:|||:|:|| .|||||||| :||| ||| |||: | |
Db        742 TTNPATLLKEAIHVISCGYEDKTEWGKEIGWIYGSVTEDIL GFKMHARGWISIYCNPPR 801

Qy        699 PAFKGSAPINLSDRLHQVLRWALGSVEIFFSRHCPLWYGFK GRLKWLQRLAYINTIVYP 758
              |||||||||||||.|:||||:|||:.:|| - ||| :|:  :|| :-:|:| ||:|:||
Db        802 PAFKGSAPINLSDRLNQVLRWALGSIEILLSRHCPIWYGY GRLRILERIAYINTIVYP 860

Qy        759 FTSLPLIAYCTIPAVCLLTGKFIIPTLSNLASMLFLGLFISIIVTAVLELRWSGVSIEDL 818
              ||:|||||| :|| ||:| :|||| :|| ||: |: ||||| || :||||||||||||
Db        861 ITSIPLIAYCILPAFCLITDRFIIPEISNYASIWFILLFISIAVTGILELRWSGVSIEDW 920

Qy        819 WRNEQFWVIGGVSAHLFAVFQGFLKMLAGIDTNFTVTAKAA-EDAEFGELYMVKWTTLLI 877
              |||||||||||| |||||||||| ||:||||||||||:||  || :| |||: ||| |||
Db        921 WRNEQFWVIGGTSAHLFAVFQGLLKVLAGIDTNFTVTSKATDEDGDFAELYIFKWTALLI 980

Qy        878 PPTTLLIINMSG-CAGFSDALNKGYEAWGPLFGKVFFAFWVILHLYPFLKGLMGRQNLTP 936
              |||||:|::|: |  || | |:| ||::|||||||:||| ||| ||||||||:|||| ||
Db        981 PPTTVLLVNLIGIVAGVSYAVNSGYQSWGPLFGKLFFALWVIAHLYPFLKGLLGRQNRTP 1040

Qy        937 TIVVLWSVLLASVFSLVWVKINPFVNKVDN 966
              |||::|||||||:|||:||:|||||:   |
Db       1041 TIVIVWSVLLASIFSLLWVRINPFVDANPN 1070
```

RESULT 4
T51579
cellulose synthase catalytic subunit (IRX3) - Arabidopsis thaliana
N;Alternate names: protein T10B6_80
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 18-Aug-2000
C;Accession: T51579
R;Sato, S.; Nakamura, Y.; Kaneko, T.; Kato, T.; Asamizu, E.; Kotani, H.; Tabata, S.; submitted to the Protein Sequence Database, August 2000
A;Reference number: Z25394
A;Accession: T51579
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1026 <SAT>
A;Cross-references: EMBL:AL391142
A;Experimental source: cultivar Columbia; BAC clone T10B6
C;Genetics:
A;Map position: 5
A;Introns: 25/3; 91/1; 184/1; 215/2; 304/2; 419/3; 507/3; 578/3; 650/1; 715/3; 833/3
A;Note: T10B6_80

Query Match 63.9%; Score 3346; DB 2; Length 1026;
Best Local Similarity 62.1%; Pred. No. 5.3e-251;
Matches 632; Conservative 116; Mismatches 167; Indels 102; Gaps 14;

```
Qy          9 CHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEINEGRKVCLRC--------GSP--- 57
              |  ||:|:|    |:|||||:|| :| |: |:|:|  || : | :|        |||
Db         37 CEICGDQIGLTVEGDLFVACNECGFPACRPCYEYERREGTQNCPQCKTRYKRLRGSPRVE 96

Qy         58 --YDENLLDDVEKKGSGNQSTMASHLNDSQDVGIHARHISSVSTVDSEMNDE-------- 107
                ||  :||:|  : :         |  |        |     :|       :||
Db         97 GDEDEEDIDDIEYEFN------IEHEQDKHKHSAEAMLYGKMSYGRGPEDDENGRFPPVI 150

Qy        108 -------YGN----------PI----------WKNRVKSCKDKENKKKKRSPKA 134
                    ||           |           |: |:   | :        |:
Db        151 AGGHSGEFPVGGGYGNGEHGLHKRVHPYPSSEAGSEGGWRERMDDWKLQHG---NLGPEP 207

Qy        135 ETEPAQVPTEQQMEEKPSAEASEPLSIVYPIPRNKLTPYRAVIIMRLVILGLFFHFRITN 194
              : :|        |     || :|||   ||  :|: ||| ||: ||||| :|  :|: |
Db        208 DDDP---------EMGLIDEARQPLSRKVPIASSKINPYRMVIVARLVILAVFLRYRLLN 258

Qy        195 PVDSAFGLWLTSVICEIWFAFSWVLDQFPKWNPVNRETYIERLSARYEREGEPSQLAGVD 254
              ||  | |||||||||||||| ||:|||||| |: ||||::||| ||||||||: || ||
Db        259 PVHDALGLWLTSVICEIWFAVSWILDQFPKWFPIERETYLDRLSLRYEREGEPNMLAPVD 318

Qy        255 FFVSTVDPLKEPPLITANTVLSILAVDYPVDKVSCYVSDDGAAMLSFESLVETAEFARKW 314
              |||||||||||||:|:||||||||:||||:|:|||||||||:||:|||| ||||||||
Db        319 VFVSTVDPLKEPPLVTSNTVLSILAMDYPVEKISCYVSDDGASMLTFESLSETAEFARKW 378

Qy        315 VPFCKKFSIEPRAPEMYFTLKVDYLQDKVHPTFVKERRAMKREYEEFKVRINAQVAKASK 374
              |||||||||||||| ||: |:|||:||| |:|||||||||:|||:|||:|| |||| |
Db        379 VPFCKKFSIEPRAPEMYFTLKVDYLQDKVHPTFVKERRAMKREYEEFKVRINAQVAKASK 438

Qy        375 TPEEGWTMQDGTPWPGNNTRDHP-------GHDSGLPWEILGARDIEGNELPRLVYVSRE 427
              | ||| |:||||||||||:|||       ||  |          |:||:|||||||||
Db        439 VPLEGWIMQDGTPWPGNNTKDHPGMIQVFLGHSGGF--------DVEGHELPRLVYVSRE 490
```

```
Qy       428 KRPGYQHHKKAGAENALVRVSAVLTNAPYILNVDCDHYVNNSKAVREAMCFLMDPQVGRD 487
Db       491 KRPGFQHHKKAGAMNALVRVAGVLTNAPFMLNLDCDHYVNNSKAVREAMCFLMDPQIGKK 550

Qy       488 VCYVQFPQRFDGIDKSDRYANRNVVFFDVNMKGLDGIQGPVYVGTGCVFNRQALYGYGPP 547
Db       551 VCYVQFPQRFDGIDTNDRYANRNTVFFDINMKGLDGIQGPVYVGTGCVFKRQALYGYEPP 610

Qy       548 SMPSLRKRKDSSSCFSCCCPSKKKPAQDPAEVYRDAKREDLNAAIFNLTEIDNYDEHERS 607
Db       611 KGP---KRPKMISCGCCPCFGRRRK--------NKKFSKNDMNGDVAALGGAEGDKEH -- 657

Qy       608 MLISQLSFEKTFGLSSVFIESTLMENGGVPESANSPPFIKEAIQVIGCGYEEKTEWGKQI 667
Db       658 -LMSEMNFEKTFGQSSIFVTSTLMEEGGVPPSSSPAVLLKEAIHVISCGYEDKTEWGTEL 716

Qy       668 GWIYGSVTEDILSGFKMHCRGWRSIYCMPVRPAFKGSAPINLSDRLNQVLRWALGSVEIF 727
Db       717 GWIYGSITEDILTGFKMHCRGWRSIYCMPKRPAFKGSAPINLSDRLNQVLRWALGSVEIF 776

Qy       728 FSRHCPLWYGFGGGRLKWLQRLAYINTIVYPFTSLPLIAYCTIPAVCLLTGKFIIPTLSN 787
Db       777 FSRHSPLWYGYKGGKLKWLERFAYANTTIYPFTSIPLLAYCILPAICLLTDKFIMPPIST 836

Qy       788 LASMLFLGLFISIIVTAVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQGFLKMLAG 847
Db       837 FASLFFISLFMSIIVTGILELRWSGVSIEEWWRNEQFWVIGGISAHLFAVVQGLLKILAG 896

Qy       848 IDTNFTVTAKAAEDAEFGELYMVKWTTLLIPPTTLLIINMSG-CAGFSDALNKGYEAWGP 906
Db       897 IDTNFTVTSKATDDDDFGELYAFKWTTLLIPPTTVLIINIVGVVAGISDAINNGYQSWGP 956

Qy       907 LFGKVFFAFWVILHLYPFLKGLMGRQNLTPTIVVLWSVLLASVFSLVWVKINPFVNK 963
Db       957 LFGKLFFSFWVIVHLYPFLKGLMGRQNRTPTIVVIWSVLLASIFSLLWVRIDPFVLK 1013
```

```
RESULT 5
S2054
ellulose synthase (EC 2.4.1.-) catalytic subunit [validated] - Arabidopsis thaliana
;Species: Arabidopsis thaliana (mouse-ear cress)
;Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 31-Dec-2003
;Accession: T52054
;Arioli, T.; Peng, L.; Betzner, A.S.; Burn, J.; Wittke, W.; Herth, W.; Camilleri, C.; H
cience 279, 717-720, 1998
;Title: Molecular analysis of cellulose biosynthesis in Arabidopsis.
;Reference number: Z13745; MUID:98111412; PMID:9445479
;Accession: T52054
;Status: preliminary; translated from GB/EMBL/DDBJ
;Molecule type: mRNA
;Residues: 1-1065 <ARI>
;Cross-references: EMBL:AF027174; PIDN:AAC39336.1
;Genetics:
;Note: Ath-B
;Function:
;Description: EC 2.4.1.-; cellulose synthase [validated, MUID:98111412]; involved in as
;Keywords: glycosyltransferase; hexosyltransferase

 Query Match             63.5%;  Score 3327;  DB 2;  Length 1065;
 Best Local Similarity   60.3%;  Pred. No. 1.7e-249;
 Matches  651;  Conservative  123;  Mismatches  188;  Indels  118;  Gaps   21;

Qy         2 MES-----GAPI-------CHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEINEGRK 49
Db         1 MESEGETAGKPMKNIVPQTCQICSDNVGKTVDGDRFVACDICSFPVCRPCYEYERKDGNQ 60

Qy        50 VCLRC--------GSP-----YDENLLDDV-------------------------EKK 69
Db        61 SCPQCKTRYKRLKGSPAIPGDKDEDGLADEGTVEFNYPQKEKISERMLGWHLTRGKGEEM 120

Qy        70 GSGNQSTMASH-----LNDSQDV-----GIHARHISSVSTV--------DSEMNDE---- 107
```

```
Db       121 GEPQYDFPVSHNGILKPHTNPQDTSIIFVAASFFELSVSSTIAGGKRLPYSSDVNQSPNRR 180

Qy       108 ---YSNFIWKNFVFSCKDKEKKEFERSFFAETEPAQVPTEQQM---- EEKPSAEA 155
Db       181 IVDPVGLGNVAWRFVDGWKMKCHENTGPVSTLAAGERGGVDIDASTDILADEALLNLEA 240

Qy       156 SEPLSIVVPIRPNPTYPAVIIMRLVILGLFFHFRITNPVDSAFGLWLTSVICEIWFAF 215
Db       241 RQPLSRKVSIPSSDIPYFRMVIMLRVIILCLFLHYRITNPVPNAFALWLVSVICEIWFAL 300

Qy       216 SWVLDQFPKWNPVNYETYIERLSARIERKGEPSQLAGVDFFVSTVDPLKEPPLITANTVL 275
Db       301 SWILDQFPKWFPVNRETYLDRLALRYDREGEPSQLAAVDIFVSTVDPLKEPPLVTANTVL 360

Qy       276 SILAVDYPVDKVSCYVSDDGAAVLSFEALVETAEFARKWVPFCKKFSIEPRAPEFYFSQK 335
Db       361 SILAVDYPVDKVSCYVSDDGAAMLSFEALVETAEFARKWVPFCKKYSIEPRAPEWYFAAK 420

Qy       336 IDYLKDKVQP--SFVKERRAMKREYEEFKVRINALVAKAQKTPEDGWTMQDGTPWPGNNTRD 395
Db       421 IDYLKDKVQP--SFVKERRAMKREYEEFKVRINALVAKAQKTPEEGWVMQDGTPWPGNNTRD 480

Qy       396 HP-----HDSILPWEILGARDIEDNELPPLVYVSREKRPGYQHHKKAGAENALVRVS 448
Db       481 HPGMIQVFLAQNGGLDTESNELPRLVYVSREKRPGFQHHKKAGAMNALVRVS 532

Qy       449 AVLTNAPYILNVDCDHYVNNSKAVREAMCFLMDPQVGRDVCYVQFPQRFDGIDKSDRYAN 508
Db       533 AVLTNGPFILNLDCDHYINNSKALREAMCFLMDPNLGKQVCYVQFPQRFDGIDKNDRYAN 592

Qy       509 RNVVFFDVNMKGLDGIQGPVYVGTGCVFNRQALYGYGPPSMPSLRKRKDSSSCFSCCCPS 568
Db       593 RNTVFFDINLRGLDGIQGPVYVGTGCVFNRTALYGYEPPIKV---KHKKPSLLSKLCGGS 649

Qy       569 KKK--PAQDPAEVYRDAKREDLNAAIFNLTEID---NYDEHERSMLISQLSFEKTFG 620
Db       650 RKKNSKAKKESDKKFGRHTDSTVPVFNLDDIEEGVEGAGFDD-EKALLMSQMSLEKRFG 708

Qy       621 LSSVFIESTLMENGGVPESANSPPFIKEAIQVIGCGYEEKTEWGKQIGWIYGSVTEDILS 680
Db       709 QSAVFVASTLMENGGVPPSATPENLLKEAIHVISCGYEDKSDWGMEIGWIYGSVTEDILT 768

Qy       681 GFKMHCRGWRSIYCMPVRPAFKGSAPINLSDRLNQVLRWALGSVEIFFSRHCPLWYGFGG 740
Db       769 GFKMHARGWRSIYCMPKLPAFKGSAPINLSDRLNQVLRWALGSVEILFSRHCPIWYGY N 827

Qy       741 GRLKWLQRLAYINTIVYPFTSLPLIAYCTIPAVCLLTGKFIIPTLSNLASMLFLGLFISI 800
Db       828 GRLKFLERFAYVNTTIYPITSIPLLMYCILPAVCLFTNQFIIPQISNIASIWFLSLFLSI 887

Qy       801 IVTAVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQGFLKMLAGIDTNFTVTAKAA- 859
Db       888 FATGILEMRWSGVSIEEWWRNEQFWVIGGVSAHLFAVFQGILKVLAGIDTNFTVTSKASD 947

Qy       860 EDAEFGELYMVKWTTLLIPPTTLLIINMSG-CAGFSDALNKGYEAWGPLFGKVFFAFWVI 918
Db       948 EDGDFAELYLFKWTTLLIPPTTLLIVNIVGVVAGVSYAINSGYQSWGPLFGKLFFAFWVI 1007

Qy       919 LHLYPFLKGLMGRQNLTPTIVVLWSVLLASVFSLVWVKINPFVNKVDNTLVAETCISIDC 978
Db      1008 VHLYPFLKGLMGRQNRTPTIVVVWSVLLASIFSLLWVRIDPFTSRVTGPDILE--CGINC 1065
```

```
RESULT 6
T08583
cellulose synthase (EC 2.4.1. ) catalytic chain - Arabidopsis thaliana
N;Alternate names: protein T22F8.250
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 11-Jun-1999 #sequence_revision 11 Jun 1999 #text_change 22-Oct-1999
C;Accession: T08583; T09014
R;Bevan, M.; Zimmermann, W.; Grueneisen, A.; Wambutt, R.; Bancroft, I.; Mewes, H.W.; May
submitted to the Protein Sequence Database, May 1999
A;Reference number: Z16442
```

```
A;Accession: T08583
A;Molecule type: DNA
A;Residues: 1-1084 <BEV>
A;Cross-references: EMBL:AL050351; GSPDB:GN00062; ATSP:T22F8.250
A;Experimental source: cultivar Columbia; BAC clone T22F8
R;Arioli, T.; Peng, L.; Betzner, A.S.; Burn, J.; Wittke, W.; Herth, W.; Camilleri, C.; H
Science 279, 717-720, 1998
A;Title: Molecular analysis of cellulose biosynthesis in Arabidopsis.
A;Reference number: Z13745; MUID:98111412; PMID:9445479
A;Accession: T09014
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1084 <ARI>
A;Cross-references: EMBL:AF027173; NID:g2827140; PIDN:AAC39335.1; PID:g2827141
A;Experimental source: cultivar Columbia
C;Genetics:
A;Gene: ATSP:T22F8.250; Ath-A
A;Map position: 4
A;Introns: 27/3; 156/3; 191/1; 254/2; 343/2; 458/3; 504/3; 546/3; 617/3; 708/1; 773/3; 8
C;Keywords: cell wall synthesis; glycosyltransferase; hexosyltransferase

  Query Match             62.5%;  Score 3274;  DB 2;  Length 1084;
  Best Local Similarity   57.5%;  Pred. No. 2.2e-245;
  Matches  623;  Conservative  154;  Mismatches  170;  Indels  136;  Gaps   23;

Qy          1 MMESGAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEINEGRKVCLRC------ 54
              : |      |  ||:::    : ||||||:||::|:|: |:|:|  || : | :|
Db         31 VQELSGQTCQICGDEIELTVSSELFVACNECAFPVCRPCYEYERREGNQACPQCKTRYKR 90

Qy         55 --GSP------YDENLLDDVEKK --- --GSGNQSTMASHLN -- --DS--------- 84
                |||       :|  :||:| :          :: ::| ||      ||
Db         91 IKGSPRVDGDDEEEEDIDDLEYEFDNGMDPEHAAEAALSSRLNTGRGGLDSAPPGSQIPL 150

Qy         85 -----QDVGIHA-RH--ISSVST --  -  ---VDSEMNDE--------------YGNP 111
                   :|  ::: ||  |   ||            ||    :                :|:
Db        151 LTYCDEDADMYSDRHALIVPPSTGYGNRVYPAPFTDSSAFPQARSMVPQKDIAEYGYGSV 210

Qy        112 IWKNRVKSCKDKENKKKK----- ----RSPKAETEPAQVPTEQQMEEKPSAEASEPLSI 161
                ||:||:  | :: :| :          |    : '   ,   |:     |  :|||
Db        211 AWKDRMEVWKRRQGEKLQVIKHEGGNNGRGSNDDDE-LDDFDMPMMD-----EGRQPLSR 264

Qy        162 VYPIPRNKLTPYRAVIIMRLVILGLFFHFRITNPVDSAFGLWLTSVICEIWFAFSWVLDQ 221
                |  ::: ||| :|: :| |||:||:|:' :||: |:||:|||| |||| ||:||
Db        265 KLPIRSSRINPYRMLILCRLAILGLFFHYRILHPVNDAYGLWLTSVICEIWFAVSWILDQ 324

Qy        222 FPKWNPVNRETYIERLSARYEREGEPSQLAGVDFFVSTVDPLKEPPLITANTVLSILAVD 281
                |||| |: ||||::||| |||:||:'| || || |||:||| .|||||||||||||||
Db        325 FPKWYPIERETYLDRLSLRYEKEGKPSGLAPVDVFVSTVDPLKEPPLITANTVLSILAVD 384

Qy        282 YPVDKVSCYVSDDGAAMLSFEALSETAEFARKWVPFCKKF--RIEPRAPEYFSQKIDYLKD 341
                |||||:|||||||||:||:|:|  :|||' .||| - |:  :|  ||||||:|:|||:
Db        385 YPVDKVACYVSDDGAAMLCFEALSETAEFARKWVPFCKKF--RIEPRAPEYFSQKMDYLKN 444

Qy        342 KVQPSFVKERRAMKRDYEEYKVRVNALVAKAQKTPEEGWTMQDGTPWPGNNTRDHP---- 397
                || |:||:|||||||||||:||:|:|||  :| || ||.|||| .|||. || ||||
Db        445 KVHPAFVRERRAMKRDYEEFKVKINALVATAQKVPEEGWTMQDGTPWPGNNVRDHPGMIQ 504

Qy        398 -- GHDSGLPWEILGARDIEGNELPRLVYVSREKRPGYQHHKKAGAENALVRVSAVLTNA 454
                 |          | || :||||:|||||||||: |||| || ::|:||||||:||
Db        505 VFLGHS--------GVRDTDGNELPRLVYVSREKRPGFDHHKKAGAMNSLIRVSAVLSNA 556

Qy        455 PYILNVDCDHYVNNSKAVREAMCFLMDPQVGRDVCYVQFPQRFDGIDKSDRYANRNVVFF 514
                \|:|||||||:|||||:||:|| :|||| |: |||||||'|||:|| |||:||||||||
Db        557 PYLLNVDCDHYINNSKAIRESMCFMMDPQSGKKVCYVQFPQRFDGIDRHDRYSNRNVVFF 616

Qy        515 DVNMKGLDGIQGPVYVGTGCVFNRQALYGYGPPSMPSLRKPKDSSSCFSC------ CCP 567
                |:||||:|||||:|||||||| |:|||:  |     :|      :|        ||
Db        617 DINMKGLDGIQGPIYVGTGCVFRRQALYGFEAP-----KKPKPPGKTCNCWPKWCCLCCG 671

Qy        568 SKKK---PAQDPAEVYRDAKREDLNAAIFNLTEIDNYDEH ----  -ERSMLISQLSFE 616
                 :||    |:|         |: :        :  ::| ||         |:    :|   |
Db        672 LRKKSKTKAKD--------KKTNTKETSKQIHALENVDEGVIVPVSNVEKRSEATQELKLE 723

Qy        617 KTFGLSSVFIESTLMENGGVPESANSPPFIKEAIQVIGCGYEEKTEWGKQIGWIYGSVTE 676
                | || | ||: | :::|||||| :|:    ::|||||| ||||:||||||:||||||||||
Db        724 KKFGQSPVFVASAVLQNGGVPRNASPACLLREAIQVISCGYEDKTEWGKEIGWIYGSVTE 783

Qy        677 DILSGFKMHCRGWRSIYCMPVRPAFKGSAPINLSDRLHQVLRWALGSVEIFFSRHCPLWY 736
                |||:|||||| ||||:|||| | ||||||||||||||||||||||||||| |||||:||
Db        784 DILTGFKMHCHGWRSVYCMPKRAAFKGSAPINLSDRLHQVLRWALGSVEIFLSRHCPIWY 843

Qy        737 GFGGGRLKWLQRLAYINTIVYPFTSLPLIAYCTIPAVCLLTGKFIIPTLSNLASMLFLGL 796
                |:||| ||||:| :|||::|||:|||||| ||::||||||||||:| :|| | :||: :
Db        844 GYGGG-LKWLERFSYINSVVYPWTSLPLIVYCSLPAVCLLTGKFIVPEISNYAGILFMLM 902

Qy        797 FISIIVTAVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQGFLKMLAGIDTNFTVTA 856
                |||| || :||::| || |:| |||||||||| |:||||:||| ||:|||::|||||:
Db        903 FISIAVTGILEMQWGGVGIDDWWRNEQFWVIGGASSHLFALFQGLLKVLAGVNTNFTVTS 962

Qy        857 KAAEDAEFGELYMVKWTTLLIPPTTLLIINMSG-CAGFSDALNKGYEAWGPLFGKVFFAF 915
                |||:|  | |||: |||||||||||||||: |   | |||:: ||::||||||:|||
Db        963 KAADDGAFSELYIFKWTTLLIPPTTLLIINIIGVIVGVSDAISNGYDSWGPLFGRLFFAL 1022

Qy        916 WVILHLYPFLKGLMGRQNLTPTIVVLWSVLLASVFSLVWVKINPFVNKVDNTLVAETCIS 975
                |||:|||||||::|:|:  |||:|:||:||||: :|:||::|||| |     | | |
Db       1023 WVIVHLYPFLKGMLGKQDKMPTIIVVWSILLASILTLLWVRVNPFVAK--GGPVLEIC-G 1079

Qy        976 IDC 978
                ::|
Db       1080 LNC 1082


RESULT 7
F84649
probable cellulose synthase catalytic subunit [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: F84649
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: F84649
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1065 <STO>
A;Cross-references: GB:AE002093; NID:g4432865; PIDN:AAD20713.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g25540
A;Map position: 2

  Query Match             62.0%;  Score 3251;  DB 2;  Length 1065;
  Best Local Similarity   59.5%;  Pred. No. 1.3e-243;
  Matches  626;  Conservative  130;  Mismatches  190;  Indels  106;  Gaps   20;

Qy          8 ICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEINEGRKVCLRC--------GSP-- 57
              ||  ||: ||    | :||||:|| :|:|:||:|:|  :| : | :|        |||
Db         32 ICQICGDDVGLTKTGNVFVACNECGFPLCQSCYEYERKDGSQCCPQCKARFRRHNGSPRV 91

Qy         58 ---YDENLLDDVEKK---GSGNQSTMASH---------------------------- 80
                    |: ::|:| :     ||     |
Db         92 EVDEKEDDVNDIENEFDYTQGNNKARLPHRAEEFSSSSRHEESLPVSLLTHGHPVSGEIP 151

Qy         81 ----------LNDSQDVGIHARHISSVSTVD-SEMNDEYG--NPIWKNRVKSCKDKENK- 126
                       | |  ||:   :  |  :| |:  : ||  |  || |::  | |::|
Db        152 TPDRNATLSPCIDPQLPGIYQLLLLPVRILDPSKDLNSYGLVNVDWKKRIQGWKLKQDKN 211

Qy        127 ------KKKRSPKAETE-PAQVPTEQQMEEKPSAEASEPLSIVYPIPRNKLTPYRAVIIM 179
                     |       | |       | || :    :|  |:| |   |  ::|||| ||::
Db        212 MIHMTGKYHEGKGGEFEGTGSNGDELQMVD----DARLPMSRVVHFPSARMTPYRIVIVL 267

Qy        180 RLVILGLFFHFRITNPVDSAFGLWLTSVICEIWFAFSWVLDQFPKWNPVNRETYIERLSA 239
                ||:|||:| |:| |:||  |: |||||||||||||||:||||||| |:||||:::||:
Db        268 RLIILGVFLHYRTTHPVKDAYALWLTSVICEIWFAFSWLLDQFPKWYPINRETFLDRLAL 327
```

```
y       240 RYEREGEPSQLAGVDFFVSTVDPLKEPPLITANTVLSILAVDYPVDKVSCYVSDDGAAML 299
o       328 RYDRDGEPSQLAPVDVFVSTVDPMKEPPLVTANTVLSILAVDYPVDKVACYVSDDGSAML 387

y       300 SFESLVETAEFARKWVPFCKKFSIEPRAPEFYFSQKIDYLKDKVQPSFVKERRAMKRDYE 359
o       388 TFEALSETAEFSKKWVPFCKKFNIEPRAPEFYFSQKIDYLKDKIQPSFVKERRAMKREYE 447

y       360 EYKVRVNALVAKAQKTPEEGWTMQDGTPWPGNNTRDHP-------GHDSGLPWEILGARD 412
o       448 EFKVRINILVAKAQKIPEDGWTMEDGTSWPGNNPRDHPGMIQVFLGHSGGL--------D 499

y       413 IEGNELPRLVYVSREKRPGYQHHKKAGAENALVRVSAVLTNAPYILNVDCDHYVNNSKAV 472
o       500 TDGNELPRLIYVSREKRPGFQHHKKAGAMNALIRVSAVLTNGAYLLNVDCDHYFNNSKAI 559

y       473 REAMCILMDPQVGRDVCYVQFPQRFDGIDKSDRYANRNVVFFDVNMKGLDGIQGPVYVGT 532
b       560 KEAMCFMMDPAIGKKCCYVQFPQRFDGIDLHDRYANRNTVFFDINLKGLDGIQGPVYVGT 619

y       533 GCVFNRQALYGYGPP-SMPSLRKRKDSSSCFSCCCPSKKKPAQDPAEVYRD---AKREDL 588
b       620 GCCFNRQALYGYDPVLTEEDLEPNIIVKSCFG----SRKKGKSRKIPNYEDNRSIKRSDS 675

y       589 NAAIFNLTEID----NYDEHERSMLISQLSFEKTFGLSSVFIESTLMENGGVPESANSPP 644
b       676 NVPLFNMEDIDEDVEGY-EDEMSLLVSQKRLEKRFGQSPVFIAATFMEQGGLPSTTNPLT 734

y       645 FIKEAIQVIGCGYEEKTEWGKQIGWIYGSVTEDILSGFKMHCRGWRSIYCMPVRPAFKGS 704
b       735 LLKEAIHVISCGYEAKTDWGKEIGWIYGSVTEDILTGFKMHARGWISIYCVPSRPAFKGS 794

y       705 APINLSDRLHQVLRWALGSVEIFFSRHCPLWYGFGGGRLKWLQRLAYINTIVYPFTSLPL 764
b       795 APINLSDRLNQVLRWALGSIEILLSRHCPIWYGY-NGRLKLLERIAYINTIVYPITSIPL 853

y       765 IAYCTIPAVCLLTGKFIIPTLSNLASMLFLGLFISIIVTAVLELRWSGVSIEDLWRNEQF 824
b       854 LAYCMLPAFCLITNTFIIPEISNLASLCFMLLFASIYASAILELKWSDVALEDWWRNEQF 913

y       825 WVIGGVSAHLFAVFQGFLKMLAGIDTNFTVTAKAA-EDAEFGELYMVKWTTLLIPPTTLL 883
b       914 WVIGGTSAHLFAVFQGLLKVFAGIDTNFTVTSKASDEDGDFAELYVFKWTSLLIPPTTIL 973

y       884 IINMSG-CAGFSDALNKGYEAWGPLFGKVFFAFWVILHLYPFLKGLMGRQNLTPTIVVLW 942
b       974 LVNLVGIVAGVSYAINSGYQSWGPLMGKLLFAFWVVAHLYPFLKGLLGRQNRTPTIVIVW 1033

y       943 SVLLASVFSLVWVKINPFVNK---VDNTLVAE 971
b      1034 SALLASIFSLLWVRINPFVSTTGVMSNSFMGE 1065
```

```
ESULT 8
84604
robable cellulose synthase catalytic subunit [imported] - Arabidopsis thaliana
;Species: Arabidopsis thaliana (mouse-ear cress)
;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
;Accession: H84604
;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
uss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
ature 402, 761-768, 1999
;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
;Reference number: A84420; MUID:20083487; PMID:10617197
;Accession: H84604
;Status: preliminary
;Molecule type: DNA
;Residues: 1-1088 <STO>
;Cross-references: GB:AE002093; NID:g4417271; PIDN:AAD20396.1; GSPDB:GN00139
;Genetics:
;Gene: At2g21770
```

```
A;Map position: 2

  Query Match           62.0%;  Score 3248;  DB 2;  Length 1088;
  Best Local Similarity  56.4%;  Pred. No. 2.4e-243;
  Matches  610;  Conservative  152;  Mismatches  187;  Indels  132;  Gaps   17;

Qy          3 ESGAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEINEGRKVCLRCGSPY---- 58
Db         33 ELSGQTCKICRDEIELTDNGEPFIACNECAFPTCRPCYEYERREGNQACPQCGTRYKRIK 92

Qy         59 ------   DENIEDVE-                 KKGSGNQSTMASHLND---     83
                    ::: : |:|                   :  :|  :   ||:
Db         93 GSPRVEGDEEDDDIMDLEHEFYGMDPEHVTEAALYYMRLNTGRGTDEVSHLYSASPGSEV 152

Qy         84 -----    -              RQVVGIHARHI -----SSVSTVDSEMNDE --   YG 109
                                     |  |          :|: |           ||
Db        153 PLLTYGTMV VHAIVVTSVCEMMLVPETTFSFASTHTRPMVPQKELTVYGYG 212

Qy        110 NPIWKNR-- ---  -EERRPSFA- EIAQL EQQMEEKPS---  ---  -- - -A 153
                 ||:      |  |  |    :::
Db        213 SVAWKDRMEVWK            KQQIEELDQVVKNERVNDGDGDGFIVDELDDPGLPMMD 261

Qy        154 EASEPLSIVYPIPRNKLTPYRAVIIMRLVILGLFFHFRITNPVDSAFGLWLTSVICEIWF 213
                |  :|||   |:  ::: ||| :|  | ||||  |:|| :|: |:|||| :|||| :
Db        262 EGRQPLSRKLPIRSSRINPYRMLIFCRLAILGLFFHYRILHPVNDAFGLWLTSVICEIWF 321

Qy        214 AFSWVLDQFPKWNPVNRETYIERLSARYEREGEPSQLAGVDFFVSTVDPLKEPPLITANT 273
                | |:||| |||  : ||||::||  ||:::|:|::| || ||||||:||||:| |||
Db        322 AVSWILDQFPKWYPIERETYLDRLSLRYEKEGKPSELAPVDVFVSTVDPLKEPPLITANT 381

Qy        274 VLSILAVDYPVDKVSCYVSDDGAAMLSFESLVETAEFARKWVPFCKKFSIEPRAPEFYFS 333
                |||||||||:|:|:|||||||||:||:| ||||||||||||||||||||||||:|
Db        382 VLSILAVDYPVEKVACYVSDDGAAMLTFEALSYTAEFARKWVPFCKKFSIEPRAPEWYFS 441

Qy        334 QKIDYLKDKVQPSFVKERRAMKRDYEEYKVRVNALVAKAQKTPEEGWTMQDGTPWPGNNT 393
                ||:|||| |  |:|  ||||:|||||:|:||::|||: :|| ||:||||| ||||:|
Db        442 QKMDYLKHKVDPAFVMERRAMKRDYEEFKVKINALVSVSQKVPEDGWTMQDGTPWPGNNV 501

Qy        394 RDHP--     -GHDSGLPWEILGARDIEGNELPRLVYVSREKRPGYQHHKKAGAENALVR 446
                ||||       :|       |  |::|||||||||||||||: ||||||  |:|::
Db        502 RDHPGMIQVFLGHS -------GVCDMDGNELPRLVYVSREKRPGFDHHKKAGAMNSLIR 553

Qy        447 VSAVLTNAPYILNVDCDHYVNNSKAVREAMCILMDPQVGRDVCYVQFPQRFDGIDKSDRY 506
                |||||:|||:|:|||||:|||:|:|:|:|:| :|||| |: :||||:|:|||||: |||
Db        554 VSAVLSNAPYLLNVDCDHYINNSKAIREAMCFMMDPQSGKKICYVQFPQRFDGIDRHDRY 613

Qy        507 ANRNVVFFDVNMKGLDGIQGPVYVGTGCVFNRQALYGYGPPSMPSLRKRKDSSSC---   561
                :||:|||:|||:|:|||||:|:|||||  |||||:: |     :|:     :|
Db        614 SNRNVVFFDINMKGLDGIQGPIYVGTGCVFRRQALYGFDAPK----KKQPPGRTCNCWPK 669

Qy        562 FSC-CCPSKKKPAQDPAEVYRDAKREDLNA- AIFNLTEIDNYDEHERSMLISQLSFEKT 618
                : | |:| : |  :  :|     :|  |: :: |       | :  :|| ||
Db        670 WCCLCCGMPKKKTGFVKDNQRKKPRETSKQIHALEHIEEGLQVTNAENNSETAQLKLEKK 729

Qy        619 FGLSSVFIESTLMENGGVPESANSPFIKEAIQVIGCGYEEKTEWGKQIGWIYGSVTEDI 678
                :| | :  :|:   ||| : |     :|||| ||||||||||:|||||:|||||
Db        730 FGQSPVLVASTLLENGGVPSNVNPASLIREAIQVISCGYEEKTEWGKEIGWIYGSVTEDI 789

Qy        679 LSGFKMHCRGWRSIYCMPVRPAFKGSAPINLSDRLHQVLRWALGSVEIFFSRHCPLWYGF 738
                :|||:||  :||:.|| : |||'':|||||||||||||||||:| |||:|:|||:
Db        790 LTGFKMHCRGWRSVYCMPKRAAFKGSAPINLSDRLHQVLRWALGSVEIFLSRHCPIWYGY 849

Qy        739 GGGRLKWLQRLAYINTIVYPFTSLPLIAYCTIPAVCLLTGKFIIPTLSNLASMLFLGLFI 798
                ||| ||||:| :|:::|:|:|||': | ::||:|||||||:| :|| | :||| :|:
Db        850 GGGKLKWLERFSYINSVVYPWTSLPLLVYCSLPAICLLTGKFIVPEISNYAGILFLLMFM 909

Qy        799 SIIVTAVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQGFLKMLAGIDTNFTVTAKA 858
                || || :|||::|  : |:| |||||||||||:|||:||| ||:||: ||||||:||
Db        909 SIAVTGILEMQWGKIGIDDWWRNEQFWVIGGVSSHLFALFQGLLKVLAGVSTNFTVTSKA 968

Qy        859 AEDAEFGELYMVKWTTLLIPPTTLLIINMSG-CAGFSDALNKGYEAWGPLFGKVFFAFWV 917
                |:| || |||: |||:|||||||||||:| |   | |||:| ||::||||||:|||: ||
Db        969 ADDGEFSELYIFKWTSLLIPPTTLLIINIVGVIVGVSDAINNGYDSWGPLFGRLFFALWV 1028
```

```
Qy        918 ILHLYPFLKGLMGRQNLTPTIVVLWSVLLASVFSLVWVKINPFVNKVDNTLVAETCISID 977
Db       1029 IVHLYPFLKGLLGKQDRVPTIILVWSILLASILTLLWVRVNPFVSK--DGPVLEIC-GLD 1085

Qy        978 C 978
Db       1086 C 1086


RESULT 9
T52028
cellulose synthase [imported] - Arabidopsis thaliana (fragment)
C;Species: Arabidopsis thaliana (mouse ear cress)
C;Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 20-Oct-2000
C;Accession: T52028
R;Coshi, C
submitted to the EMBL Data Library, May 1998
A;Reference number: Z25890
A;Accession: T52028
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1081 <JOS>
A;Cross-references: EMBL:AF062485; PIDN:AAC29067.1

  Query Match             61.1%;  Score 3199.5;  DB 2;  Length 1081;
  Best Local Similarity   56.3%;  Pred. No. 1.4e-239;
  Matches  607;  Conservative  151;  Mismatches  195;  Indels  125;  Gaps   18;

Qy          1 MMESGAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEINEGRKVCLRC------ 54
Db         28 VQELSGQTCQICRDEIELTVDGEPFVACNECAFPVCRPCYEYERREGNQACPQCKTRFKR 87

Qy         55   GSP ----YDENLLDD-------------------VEKKGSG-NQSTMASHLNDS 84
Db         88 LKGSPRVEGDEEEDDIDDLDNEFEYGNNGIGFDQVSEGMSISRRNSGFPQSDLDSAPPGS 147

Qy         85 Q-------------------   DVGIHARHISSVSTVDMEMNDE--------   - 107
Db        148 QIPLLTYGDEDVEISSDRHALIVPPSLGGHGNRVHPVSLSDPTVAAHRRLMVPQKDLAVY 207

Qy        108 -YGNPIWKNRVKSCKDKENKKKKRSPKAETEPACVPTEQCMEEKPSA------EASEPLS 160
Db        208 GYGSVAWKDRMEEWKRKQN-EKLQVVRHEGDP  -----DFEDGDDADFPMMDEGRQPLS 259

Qy        161 IVYPIPRNKLTPYRAVIIMRLVILSLFFHFRITNPVDSAFGLWLTSVICEIWFAFSWVLD 220
Db        260 MKIPIKSSKINPYRMLIVLRLVILSLFFHYRILHPVKDAYALWLISVICEIWFAVSWVLD 319

Qy        221 QFPKWNPVNRETYIERLSARYEREGEPSQLAQVDFFVSTVDPLKEPPLITANTVLSILAV 280
Db        320 QFPKWYPIERETYLDRLSLRYEKEGKPSGLSPVDVFVSTVDPLKEPPLITANTVLSILAV 379

Qy        281 DYPVDKVSCYVSDDGAAMLSFESLVETAEFARKWVPFCKKFSIEPRAPEFYFSQKIDYLK 340
Db        380 DYPVDKVACYVSDDGAAMLTFEALSETAEFARKWVPFCKKYSIEPRAPEWYFCHKMDYLK 439

Qy        341 DKVQPSFVKERRAMKRDYEEYKVRVNALVAKAQKTPEEGWTMQDGTPWPGNNTRDHP--- 397
Db        440 NKVHPAFVRERRAMKRDYEEFKVKINALVATAQKVPEDGWTMQDGTPWPGNSVRDHPGMI 499

Qy        398 --  GHDSGLPWEILGARDIEGNELPRLVYVSREKRPGYQHHKKAGAENALVRVSAVLTN 453
Db        500 QVFLGSD--------GVRDVENNELPRLVYVSREKRPGFDHHKKAGAMNSLTRVSGVLSN 551

Qy        454 APYILNVDCDHYVNNSKAVREAMCILMDPQVGRDVCYVQFPQRFDGIDKSDRYANRNVVF 513
Db        552 APYLLNVDCDHYINNSKALREAMCFMMDPQSGKKICYVQFPQRFDGIDRHDRYSNRNVVF 611

Qy        514 FDVNMKGLDGIQGPVYVGTGCVFNRQALYGYGPPSMPSLRKRKDSSSCFSC-------CC 566
Db        612 FDINMKGLDGLQGPIYVGTGCVFRRQALYGFDAP   --KKKKGPRKTCNCWPKWCLLCF 666

Qy        567 PSKKKPAQDPAEVYRDAKREDLNAAIFNLTEIDNYDEH-----ERSMLISQLSFEKTFGL 621
Db        667 GSRKNRKAKTVAADKKKKNREASKQIHALENIEEGRGHKVLNVEQSTEAMQMKLQKKYGQ 726

Qy        622 SSVFIESTLMENGGVPESANSPPFIKEAIQVIGCGYEEKTEWGKQIGWIYGSVTEDILSG 681
Db        727 SPVFVASARLENGGMARNASPACLLKEAIQVISRGYEDKTEWGKEIGWIYGSVTEDILTG 786

Qy        682 FKMHCRGWRSIYCMPVRPAFKGSAPINLSDRLHQVLRWALGSVEIFFSRHCPLWYGFGGG 741
Db        787 SKMHSHGWRHVYCTPKLAAFKGSAPINLSDRLHQVLRWALGSVEIFLSRHCPIWYGYGGG 846

Qy        742 RLKWLQRLAYINTIVYPFTSLPLIAYCTIPAVCLLTGKFIIPTLSNLASMLFLGLFISII 801
Db        847 -LKWLERLSYINSVVYPWTSLPLIVYCSLPAICLLTGKFIVPEISNYASILFMALFSSIA 905

Qy        802 VTAVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQGFLKMLAGIDTNFTVTAKAAED 861
Db        906 ITGILEMQWGKVGIDDWWRNEQFWVIGGVSAHLFALFQGLLKVLAGVDTNFTVTSKAADD 965

Qy        862 AEFGELYMVKWTTLLIPPTTLLIINMSG-CAGFSDALNKGYEAWGPLFGKVFFAFWVILH 920
Db        966 GEFSDLYLFKWTSLLIPPMTLLIINVIGVIVGVSDAISNGYDSWGPLFGRLFFALWVIIH 1025

Qy        921 LYPFLKGLMGRQNLTPTIVVLWSVLLASVFSLVWVKINPFVNKVDNTLVAETCISIDC 978
Db       1026 LYPFLKGLLGKQDRMPTIIVVWSILLASILTLLWVRVNPFVAK--GGPILEIC-GLDC 1080


RESULT 10
T10800
cellulose synthase (EC 2.4.1.-) catalytic chain celA2 - upland cotton (fragment)
C;Species: Gossypium hirsutum (upland cotton)
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
C;Accession: T10800
R;Pear, J.R.; Kawagoe, Y.; Schreckengost, W.E.; Delmer, D.P.; Stalker, D.M.
Proc. Natl. Acad. Sci. U.S.A. 93, 12637-12642, 1996
A;Title: Higher plants contain homologs of the bacterial celA genes encoding the cata
A;Reference number: Z17152; MUID:97057296; PMID:8901635
A;Accession: T10800
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-685 <PEA>
A;Cross-references: EMBL:U58284; NID:g1706957; PIDN:AAB37767.1; PID:g1706958
A;Experimental source: strain Acala SJ-2; fiber
C;Genetics:
A;Gene: celA2
C;Function:
A;Description: involved in synthesis of cellulose
C;Keywords: glycosyltransferase; hexosyltransferase

  Query Match             52.1%;  Score 2730;  DB 2;  Length 685;
  Best Local Similarity   73.7%;  Pred. No. 1.9e-203;
  Matches  509;  Conservative   71;  Mismatches   81;  Indels   30;  Gaps    9;

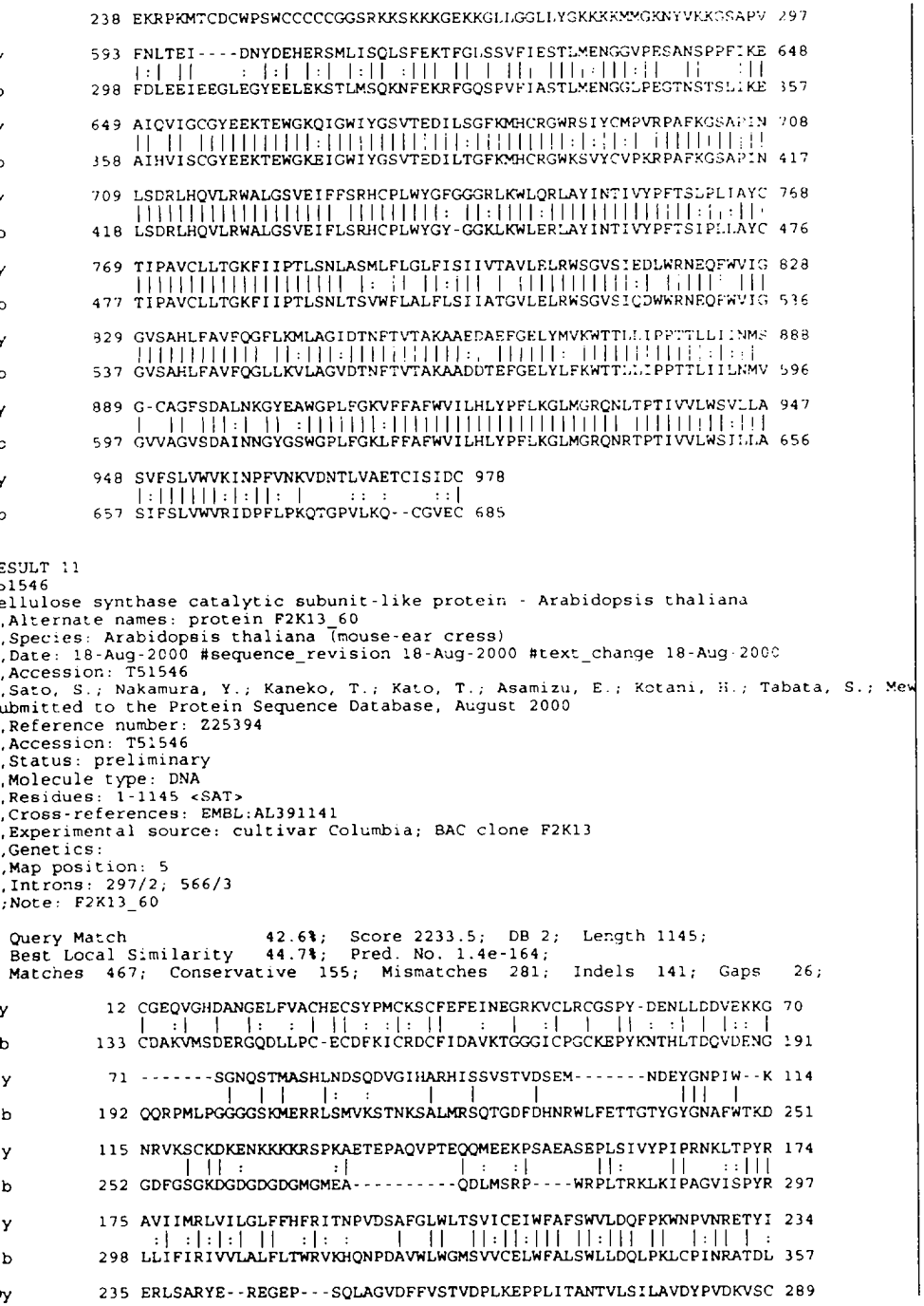
Qy        312 RKWVPFCKKFSIEPRAPEFYFSQKIDYLKDKVQPSFVKERRAMKRDYEEYKVRVNALVAK 371
Db          1 RRWVPFCKKHNVEPRAPEFYFNEKIDYLKDKVHPSFVKERRAMKREYEEFKVRINALVAK 60

Qy        372 AQKTPEEGWTMQDGTPWPGNNTRDHPGHDSGLPWEILGARDIEGNELPRLVYVSREKRPG 431
Db         61 AQKKPEEGWVMQDGTPWPGNNTRDHPGMIQ-VYLGSAGALDVDGKELPRLVYVSREKRPG 119

Qy        432 YQHHKKAGAENALVRVSAVLTNAPYILNVDCDHYVNNSKAVREAMCILMDPQVGRDVCYV 491
Db        120 YQHHKKAGAENALVRVSAVLTNAPFILNLDCDHYINNSKAMREAMCFLMDPQFGKKLCYV 179

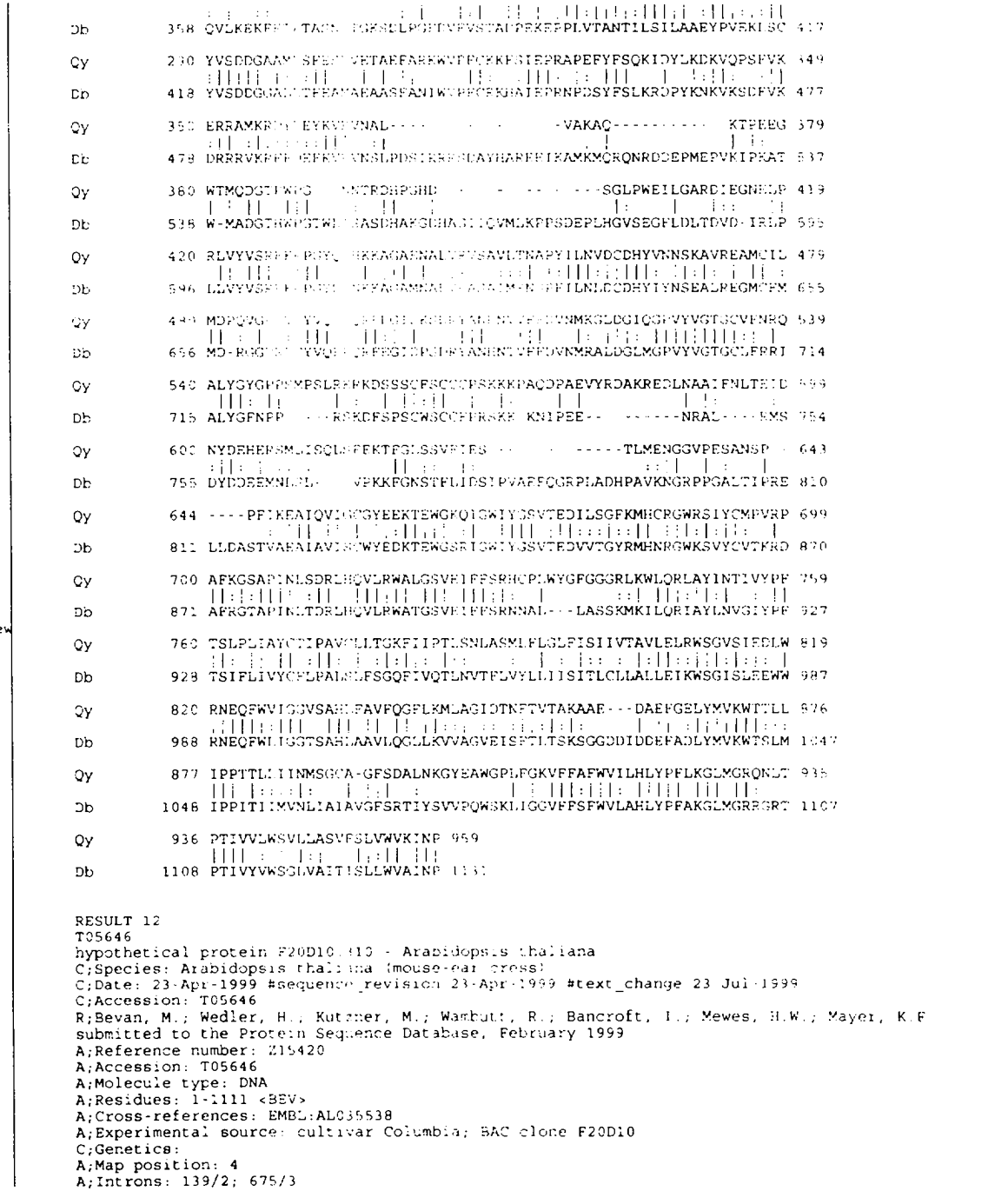
Qy        492 QFPQRFDGIDKSDRYANRNVVFFDVNMKGLDGIQGPVYVGTGCVFNRQALYGYGPPSMPS 551
Db        180 QFPQRFDGIDRHDRYANRNVVFFDINMLGLDGLQGPVYVGTGCVFNRQALYGYDPP--VS 237

Qy        552 LRKRKDSSSCF----SCCCPSKKKPAQDPAE--------VYRDAKR-------EDLNAAI 592
```



```
ESULT 11
>1546
ellulose synthase catalytic subunit-like protein - Arabidopsis thaliana
,Alternate names: protein F2K13_60
,Species: Arabidopsis thaliana (mouse-ear cress)
,Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 18-Aug-2000
,Accession: T51546
,Sato, S.; Nakamura, Y.; Kaneko, T.; Kato, T.; Asamizu, E.; Kotani, H.; Tabata, S.; Mew
ubmitted to the Protein Sequence Database, August 2000
,Reference number: Z25394
,Accession: T51546
,Status: preliminary
,Molecule type: DNA
,Residues: 1-1145 <SAT>
,Cross-references: EMBL:AL391141
,Experimental source: cultivar Columbia; BAC clone F2K13
,Genetics:
,Map position: 5
,Introns: 297/2; 566/3
;Note: F2K13_60

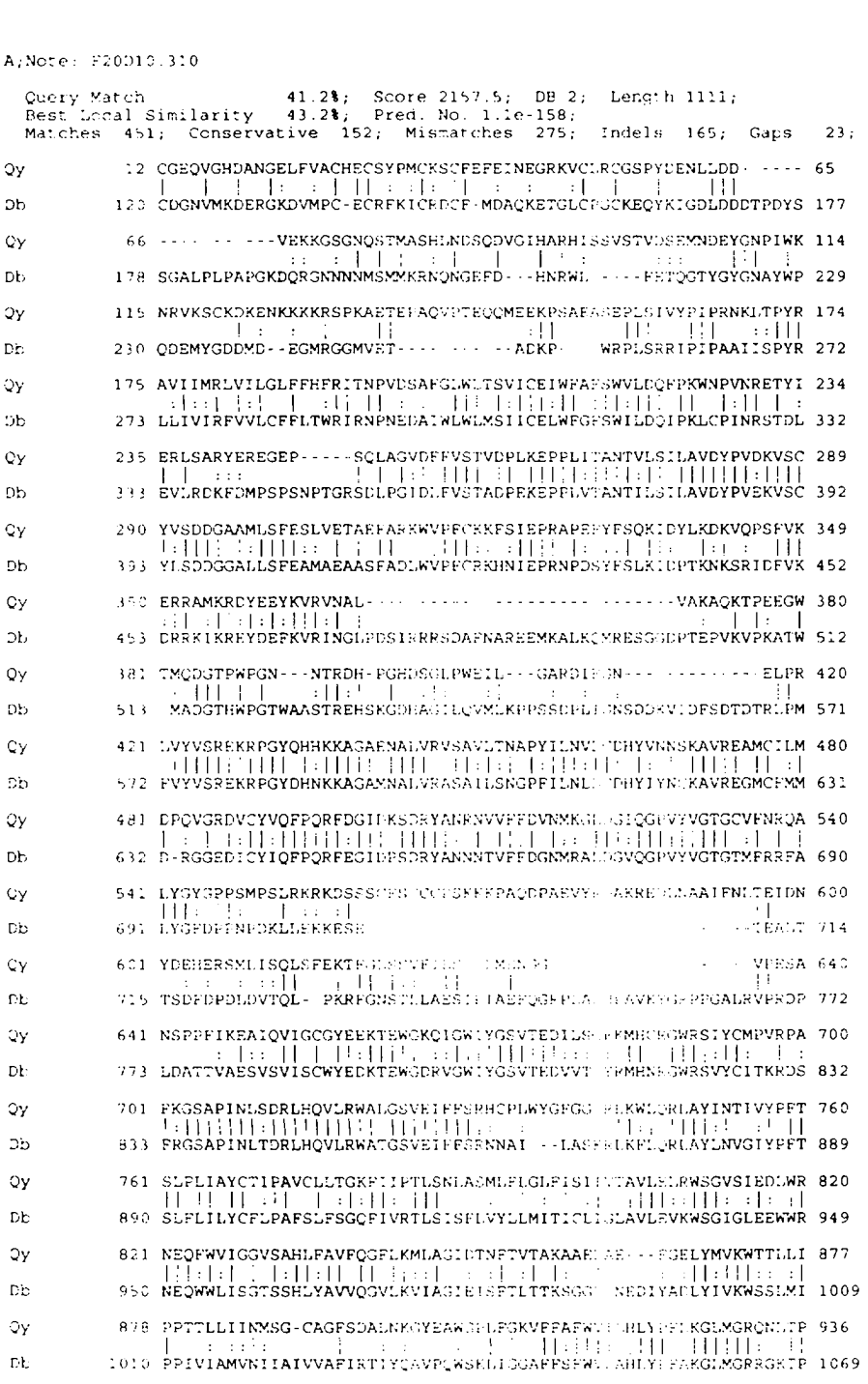
 Query Match             42.6%;  Score 2233.5;  DB 2;  Length 1145;
 Best Local Similarity   44.7%;  Pred. No. 1.4e-164;
 Matches  467;  Conservative  155;  Mismatches  281;  Indels  141;  Gaps   26;
```

```
RESULT 12
T05646
hypothetical protein F20D10.410 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 23 Jul-1999
C;Accession: T05646
R;Bevan, M.; Wedler, H.; Kutzner, M.; Wambutt, R.; Bancroft, I.; Mewes, H.W.; Mayer, K.F
submitted to the Protein Sequence Database, February 1999
A;Reference number: Z15420
A;Accession: T05646
A;Molecule type: DNA
A;Residues: 1-1111 <BEV>
A;Cross-references: EMBL:AL035538
A;Experimental source: cultivar Columbia; BAC clone F20D10
C;Genetics:
A;Map position: 4
A;Introns: 139/2; 675/3
```

A;Note: F20D10.310

Query Match 41.2%; Score 2157.5; DB 2; Length 1111;
Best Local Similarity 43.2%; Pred. No. 1.1e-158;
Matches 451; Conservative 152; Mismatches 275; Indels 165; Gaps 23;



Qy 937 TIVVLWSVLLASVFSLVWVKINP 959
Db 1070 TIVFVWAGLIAITISLLWTAINP 1092

RESULT 13
D86157
hypothetical protein F22D16.26 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Dec-2001
C;Accession: D86157
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: D86157
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1181 <STO>
A;Cross-references: GB:AE005172; NID:g6056428; PIDN:AAF02892.1; GSPDB:GN00141
C;Genetics:
A;Map position: 1

Query Match 40.9%; Score 2142; DB 2; Length 1181;
Best Local Similarity 42.5%; Pred. No. 1.9e-157;
Matches 458; Conservative 163; Mismatches 264; Indels 192; Gaps 29;

```
        663 CDHYIYNSMALREGMCFMLD-RGGDRICYVQFPQRFEGIDPNDRYANHNTVFFDVSMRAL 721

        522 DGIQGPVYVGTGCVFNRQALYGYGPPSMPS----LRKRKDSSSCFSCCCPSKKKPAQDPA 577
            ||:|||:||||||:| | ||||: ||        | :||              |   : |
        722 DGLQGPMYVGTGCIFRRTALYGFSPPRATEHHGWLGRRK----------VKISLRRPK 769

        578 EVYRDAKREDLNAAI---FNLTEIDNYDEHERSMLISQLSFEKTFGLSSVFIES-- --- 628
            : :  | ::::  |   :|  | |: |       |  |    | || |: |: | :
        770 AMMK--KDDEVSLPINGEYNEEENDDGD-------IESLLLPKRFGNSNSFVASIPVAEY 820

        629 ------TLMENG---------GVPESANSPPFIKEAIQVIGCGYEEKTEWGKQIGWIYGS 673
                  |   |          ||        : ||| || | ||:||||:|::|||||
        821 QGRLIQDLQGKGKNSRPAGSLAVPREPLDAATVAEAISVISCFYEDKTEWGKRVGWIYGS 880

        674 VTEDILSGFKMHCRGWRSIYCMPVRPAFKGSAPINLSDRLHQVLRWALGSVEIFFSRHCP 733
            ||||:::|::|| |||||||:  | ||:|:|||||:||||||||; |||:||||':
    )   881 VTEDVVTGYRMHNRGWRSIYCVTKRDAFRGTAPINLTDRLHQVLRWATGSVEIFFSRNNA 940

    /   734 LWYGFGGGRLKWLQRLAYINTIVYPFTSLPLIAYCTIPAVCLLTGKFIIPTLSNLASMLF 793
            :   |   |:|:|||:|| |  :|||||| || || :|': | :|:|: :|     :
    )   941 I---FATRRMKFLQRVAYFNVGMYPFTSLFLIVYCILPAISLFSGQFIVQSLDITFLIYL 997

    /   794 LGLFISIIVTAVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQGFLKMLAGIDTNFT 853
            | : ::: : ::||::|||::: : ||||||||||| |||  || || ||::||:| :||
    )   998 LSITLTLCMLSLLEIKWSGITLHEWWRNEQFWVIGGTSAHPAAVLQGLLKVIAGVDISFT 1057

    /   854 VTAKAAE----DAEFGELYMVKWTTLLIPPTTLLIINMSGCA-GFSDALNKGYEAWGPLF 908
            :|:|::     | || :||:|||: |::|| |::::||   | | :  |   :  :  |
    )  1058 LTSKSSAPEDGDDEFADLYVVKWSFLMVPPLTIMMVNMIAIAVGLARTLYSPFPQWSKLV 1117

    /   909 GKVFFAFWVILHLYPFLKGLMGRQNLTPTIVVLWSVLLASVFSLVWVKINPFVNKVD 965
            | |||:|||: ||||| ||||||:   |||| :|| ||: : ||:|| ,||   | |
    )  1118 GGVFFSFWVLCHLYPFAKGLMGRRGRVPTIVFVWSGLLSIIVSLLWVYINPPSGKQD 1174
```

```
ESULT 14
84741
robable cellulose synthase [imported] - Arabidopsis thaliana
;Species: Arabidopsis thaliana (mouse-ear cress)
;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
;Accession: D84741
;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
 ; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
uss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
ature 402, 761-768, 1999
;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
;Reference number: A84420; MUID:20083487; PMID:10617197
;Accession: D84741
;Status: preliminary
;Molecule type: DNA
;Residues: 1-1036 <STO>
;Cross-references: GB:AE002093; NID:g2924781; PIDN:AAC04910.1; GSPDB:GN00139
;Genetics:
;Gene: At2g33100
;Map position: 2

 Query Match             39.7%;  Score 2080.5;  DB 2;  Length 1036;
 Best Local Similarity   43.0%;  Pred. No. 9e-153;
 Matches  440;  Conservative  160;  Mismatches  256;  Indels  167;  Gaps   25;

y         64 DDVEKKGSGNQSTMASHLN---------DSQDVGIHARHISSVSTVDSEMN--------- 105
             ||::   ||: |    ::|         |:|  |      || || :|: :
b         42 DDMDV--SGDYSGQNDYINYTVLMPPTPDNQPAG------SSGSTSESKGDANRGGGGGD 93

)y       106 -DEYGNPIWKNRVKSCKDKENKKKKRSPKAETEPAQVPTEQQ----------MEEKPSAE 154
               : || : : |:   |          ||   : :  :   | :             ||  : :
b         94 GPKMGNKL-ERRLSVMKSNNKSMLLRSQTGDFDHNRWLFESKGKYGIGNAFWSEEDDTYD 152

y        155 AS----------EPLSIVYPIPRNKLTPYRAVIIMRLVILGLFFHFRITNPVDSAFGLW 203
                        :||:    ||   |:||| :|::|||:  |  :|||||  : |  ||
b        153 GGVSKSDFLDKPWKPLTRKVQIPAKILSPYRLLIVIRLVIVFFFLWWRITNPNEDAMWLW 212

y        204 LTSVICEIWFAFSWVLDQFPKWNPVNRETYIERLSARYEREGEP-----SQLAGVDFFVS 258
```

```
                |::|:|||||:|:|   || |::|| . :  :  ::|:          | | |:| |||
Db       213 GLSIVCEIWFAFSWILDILPKLNPINRATDLAALHDKFEQPSPSNPTGRSDLPGVDVFVS 272

Qy       259 TVDPLKEPPLITANTVLSILAVDYPVDKVSCYVSDDGAAMLSFESLVETAEFARKWVPFC 318
             | || ||| |:| :::|||:|||,|::|:  :|| | |:|:||:: |   || |:|||
Db       273 TADPEKEPPLVTANTLLSILAVDYPIEKLSAYISDDGGAILTFEAMAEAVRFAEYWVPFC 332

Qy       319 KKFSIEPRAPEFYFSQKIDYLKDKVQPSFVKERRAMKRDYEEYKVRVNALVAKAQKTPEE 378
             :|  |||  :: |,  | ,  |:|  :  |,:|| ::|:|:|:|||:| |  : :|  |:
Db       333 RKHDIEPRNPDSYFSLKKDPTKNKKRQDFVKDRRWIKREYDEFKVRINGLPEQIKKRAEQ 392

Qy       379 ------  - --        ------       GWTMQDGTPWPGNNTRDHPGHDSGLPWEI 407
                                              : | ||| |||       | |  ,    |
Db       393 PNMREELKEFRIAW FNGGVLFPDGVEVVKATW MADGTHWPGTWFEPKPDHSKGDHAGI 452

Qy       408 L--------          -- GAPDIENNFL  PRLVYVSREKRPGYQHHKKAGAENALVRV 447
             |                    ||    | ::  |   ||||:||: ::|||: | : :
Db       453 LQIMSFW-I HPCM  PNFSALFFGTLIGPVFMPAYVSREKRPGFDHNKKAGAMNGMVRA 511

Qy       448 SAVLTNG ' L--   ---  ---  ---  ---  ---  ---  ---  ---  ---  --- 507
             ||: :|  : |||:   ||  ||  :  |  :  :  :|:|||:.:|:|| '||||
Db       512 SAILSNGAFILNGS EHYIYNGKAIKGLMFFMMD PGGDRICYIQFPQRFEGIDPSDRYA 570

Qy       508 NRNVVFFDVNMKGLDGIQGPVYVGTGCVFNRQALYGYGPPSMPSLRKGKDSSSCFSCTCP 567
             : | ||:| ||: ||||:|||||,|||:  |  :||: ||      : | |  |
Db       571 NHNTVFFDGNMRALDGLQGPVYVGTGCMFRRYALYGFNPP------RANEYSGVFG---- 620

Qy       568 SKKKPA   QDPAEVYRDAKREDLNAAIFNLTEIDNYDEHERSMLISQLSFEKTFGLSSV 624
             :| ||  :  :::  :  ::  ||  :    |:  ! !            |  | :  |::
Db       621 QEKAPAMHVFTQSLARQTSQASDLESD   TQPLNDD--------PDLGLPKKPGNSTM 668

Qy       625 FIES--     ---    --TLMENGGVPESANSP------PFIKEAIQVIGCGYEEKTEWG 664
             | ::                 ::||  : :       | : |||  || | ':|: |:|'
Db       669 FTDTIPVAEYQGRPLADHMSVKNGRPPGALLLPRPPLDAPTVAEAIAVISCWYEDNTEWG 728

Qy       665 KQIGWIYGSVTEDILSGFKMHCRGWRSIYCMPVRPAFKGSAPINLSDRLHQVLRWALGSV 724
             :||||:|||':  :  :  ||:|: :|||| :||:  : ||:|:|||||:|||||| || |:|
Db       729 DRIGWIYGSVTEDVVTGYRMHNRGWRSVYCITKRDAFRGTAPINLTDRLHQVLRWATGSV 788

Qy       725 EIFFSRHCPLWYGFGGRLKWLQRLAYINTIVYPFTSLPLIAYCTIPAVCLLTGKFIIPT 784
             |||||:.  :   :  ||:|: :|||: |: |: || :||:|| :||| :
Db       789 EIFFSKNNAM---FATRRLKFLQRVAYLNVGIYPFTSIFLVVYCFLPALCLFSGKFIVQS 845

Qy       785 LSNLASMLFLGLFISIIVT--- AVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQG 840
             |     :'  :  :| ||    :: |:|:|| :  :: |||||||:||| |||| |  |
Db       846 L-  -DIHFLSYLVFITVTVTLISLLEVKWSGIGLEEWWRNEQFWLIGGTSAHLAAVVQG 901

Qy       841 FLKMLAGIDTNFTVTAKAA --EDAEFGELYMVKWTTLLIPPTTLLIINMSGCA-GFSDA 896
             ||::'||: :||:::||:   ||  | :||:'||| | | | |::|:|:     | :
Db       902 LLKVIAGISISFTLTSKASGEDEDDIFADLYIVKWTGLFIMPLTIIIVNLVAIVIGASRT 961

Qy       897 LNKGYEAWGPLFGKVFFAFWVILHLYPFLKGLMGRQNLTPTIVVLWSVLLASVFSLVWVK 956
             :      :| |   :||: ||: |:|:| ||| ||:    |:|| :|' |:    ||:|:
Db       962 IYSVIPQWGKLMGGIFFSLWVLTHMYPFAKGLMGRRGKVPTIVYVWSGLVSITVSLLWIT 1021

Qy       957 INP 959
             |:|
Db      1022 ISP 1024
```

```
RESULT 15
C86446
probable cellulose synthase catalytic subunit [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: C86446
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
```

A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: C86446
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-979 <STO>
A;Cross-references: GB:AE005172; NID:g10801364; PIDN:AAF23456.1; GSPDB:GN00141
C;Genetics:
A;Map position: 1

Query Match 39.3%; Score 2059.5; DB 2; Length 979;
Best Local Similarity 45.3%; Pred. No. 3.5e-151;
Matches 435; Conservative 138; Mismatches 245; Indels 143; Gaps 22;

```
Qy         96 SVSTVDSEMNDEYGNPIWKNRVKSCKDKENKKKKKRSPKAETSPAQVPTEQQMEEKPSAEA 155
              |:  :: : | |    :  : |      :|         | :   |        ||:
Db         50 SIVLLEIDSNQESVPSVSGDIVSGSSGKDN---------EIDLTDVRINVGEEEEDDTLL 100

Qy        156 SEPLSIVYPIPR-NKLTP-----YRAVIIMRLVILGLFFHFRITNPVDSAFGLWLTSVIC 209
              |:   ' | : |  |::|     || :|::|:| | ||  :|| || : |  ||| ||||
Db        101 SK---ISYSLTRVVKISPIIIALYRILIVVRVVSLALFLFWRIRNPNNKALWLWLLSVIC 157

Qy        210 EIWFAFSWVLDQFPKWNPVNRETYIERLSARYEREGE--   PSQLAGVDFFVSTVDPLK 264
              |:|'|||:||| ||  |||  ! || | | :|          | | |:| |||| |  |
Db        158 ELWFAFSWILDQIPKLFPVNHATDIEALKATFETPNPDNPTGKSDLPGIDVFVSTADAEK 217

Qy        265 EPPLITANTVLSILAVDYPVDKVSCYVSDDGAAMLSFESLVETAEFARKWVPFCKKFSIE 324
              |||||||||:||||:|||||:|:| |:|||| ::::||:: | | ||: |||||:|  ||
Db        218 EPPLVTANTILSILSVDYPVEKLSVYISDDGGSLVTFEAIAEAASFAKIWVPFCRKHKIE 277

Qy        325 PRAPEFYFSQKIDYLKDKVQPSFVKERRAMKRDYEEYKVRVNALVAKAQK---------- 374
              || || ||  | |  |||||:  :|:|,| .|| ::|:'|||||    ::
Db        278 PRNPESYFGLKRDPYKDKVRIDFVRERRYVKRAYDEFKVRVNALPHSIRRRSDAFNSKEE 337

Qy        375 --------------------------TPEEGWTMQDGTPWPGNNTRDHPGHDSG----- 402
                                       ::  | | ||| |||      | |  |
Db        338 IKALEKWKHWKVKVEEDQIKEPRPALVAPKATW-MSDGTHWPGTWAVSGPHHSRGDHASV 396

Qy        403 ----------LPWEILG----ARDIEGNE--LPRLVYVSREKRPGYQHHKKAGAENALVR 446
                       | | |    | :|| :  || ||||| |||| |:|||| |||||
Db        397 IQVLLDPPGDEPVEGKGGEGRALDLEGVDIRLPMLVYVSREKRPGYDHNKKAGAMNALVR 456

Qy        447 VSAVLTNAPYILNVDCDHYVNNSKAVREAMCILMDPQVGRFVCYVQFPQRFDGIDKSDRY 506
              ||:::| |:|||:|||||:| ||:| |: :  :||   |    ||| ||||:||| ||||
Db        457 ASAIMSNGPFILNLDCDHYVYNSRAFRDGICFMMDHE-GDRVSYVQFPQRFEGIDPSDRY 515

Qy        507 ANRNVVFFDVNMKGLDGIQGPVYVGTGCVFNRQALYGYGP  MPSIKKPKDSSSCFSCCC 566
              ||:| :.||:|:: |||| '|:| .'||:    ||   |          | |
Db        516 ANKNTVFFDINLRALDGIQGPMYVGTGSFFRTALYGFDP  VFVK-EMPSGSYCF  -- 571

Qy        567 PSKKKPAQDPAEVYRDAKRDDLNAAIFNCIEIDNYDEHEK YLINQLSFEKTFGLSSVFI 626
              |  |:  : || |                :| .  : |    |  :   | :| ||: :
Db        572 PLIKK--RSPATV ------        -ASEPEYYTDEEDE DIGLI  RKQFGSSSMLV 612

Qy        627 EST-LMENGGVP------ESANSPP------------FIEEAIQVISCGYEEKTEWGKQ 666
              |  : |  | |           ||                :' .  !! . ||:|||||
Db        613 NSVKVAEFEGRPLATVHSSRLGRPPGSLTGSRKPLDFATVNEAVNVISCWYEDKTEWGFN 672

Qy        667 IGWIYGSVTEDILSGFKMHCRGWRSIYCMPVRPAFKGSAPINLSDRLHQVLRWALGSVEI 726
              :|||||||||:::||:|| :|| | | :    ::||' |||:|||:|||| |||||
Db        673 VGWIYGSVTEDVVTGFRMHEKGWRSFYCVTEPDAFRGSAPINLTDRLHQVLRWATGSVEI 732

Qy        727 FFSRHCPLWYGFGGGRLKWLQRLAYINTIVYPFTSLPLIAYCTIPAVCLLTGKFIIPTLS 786
              ||:|:  :   | | :|| |||:||:|  :|||:|: :: || : : ! :| |:: ||:
Db        733 FFSRNNAI---FAGPKLKLLQRIAYLNVGIYPFTSIFILTYCFLPPLSLFSGHFVVETLT 789

Qy        787 NLASMLFLGLFISIIVTAVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQGFLKMLA 846
                :  | : :|:   |||:::||:|:|: ||'||||:|| |||| || || ||::!
Db        790 GSFLIYLLIITLSLCGLAVLEVKWSGISLEEWWRNEQFWLIGGTSAHLVAVLQGILKVIA 849

Qy        847 GIDTNFTVTAKAA-----EDAEFGELYMVKWTTLLIPPTTILIINM-----SGCAGFSDA 896
```

```
              |:: :||:|:|::     || || :||: ||| |:||| |::|:|:     : |     :
Db        850 GVEISFTLTSKSSTGGDDEDDEFADLYLFKWTALMIPPLTIIILNIVAILFAVCRTVFSA 909

Qy        897 LNKGYEAWGPLFGKVFFAFWVILHLYPFLKGLMGRQNLTPTIVVLWSVLLASVFSLVWVK 956
                    |  | |  ||| ||:|:|:||| ||||||   |||:| :|| |:|   ||:::
Db        910  ---NPQWSNLLGGTFFASWVLLHMYPFAKGLMGRGGKTPTVVYVWSGLIAICLSLLYIT 965

Qy        957 I 957
              |
Db        966 I 966
```

Search completed: October 29, 2003, 18:31:28
Job time : 38.2128 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 29, 2003, 18:25:58 ; Search time 3.89842 Seconds
(without alignments)
3705.922 Million cell updates/sec

Title: US-09-980-043-4
Perfect score: 324
Sequence: 1 MMESGAPICHTCGEQVGHDA.....KSCFEFEIKEGRKVCLRCGS 56

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_23:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score obtained in a search with a database of the same size, and is derived by analysis of the distribution of scores.

SUMMARIES

Result No	Score	Query Match	Length	DB	ID	Description
1	324	100.0	56	10	Q9LLJ2	Q9llj2 populus tre
2	319	98.5	978	10	O81368	O81368 populus tre
3	264	81.5	974	10	P93155	P93155 gossypium h
4	264	81.5	974	10	Q8W1W0	Q8w1w0 gossypium h
5	247	76.2	985	10	Q9C5Z8	Q9c5z8 arabidopsis
6	246	75.9	958	10	Q9SN37	Q9sn37 arabidopsis
7	246	75.9	985	10	Q8LPK5	Q8lpk5 arabidopsis
8	220	67.9	205	10	Q9SYX4	Q9syx4 gossypium h
9	220	67.9	205	10	Q9XFC6	Q9xfc6 gossypium r
10	220	67.9	205	10	Q9XFC5	Q9xfc5 gossypium r
11	220	67.9	205	10	Q9XFC4	Q9xfc4 gossypium h
12	211	65.1	205	10	Q9SYX5	Q9syx5 gossypium h
13	193.5	59.7	978	10	Q9AXK0	Q9axk0 zinnia elleg
14	192	59.3	197	10	Q93WX2	Q93wx2 gossypium s
15	192	59.3	197	10	Q8GV66	Q8gv66 gossypium d
16	192	59.3	197	10	Q8GV65	Q8gv65 gossypium s
17	192	59.3	197	10	Q8GV64	Q8gv64 gossypium t
18	192	59.3	197	10	Q8GV63	Q8gv63 gossypium t
19	191	59.0	197	10	Q8GV67	Q8gv67 gossypium g
20	189	58.3	193	10	Q9M693	Q9m693 gossypioide
21	189	58.3	197	10	Q93WX1	Q93wx1 gossypium l
22	189	58.3	197	10	Q93WX0	Q93wx0 gossypium b
23	188	58.0	197	10	Q93WX3	Q93wx3 gossypium a
24	187	57.7	192	10	Q94JP7	Q94jp7 thespesia l
25	187	57.7	192	10	Q94JP6	Q94jp6 thespesia t
26	187	57.7	192	10	Q94JP8	Q94jp8 hampea appe
27	187	57.7	193	10	Q94JP3	Q94jp3 cienfuegosi
28	184	56.8	193	10	Q94JP5	Q94jp5 thespesia p
29	183	56.5	192	10	Q94JP4	Q94jp4 lebronnecia
30	183	56.5	193	10	Q94JP9	Q94jp9 kokia dryna
31	181	55.9	194	10	Q94JP2	Q94jp2 malva sylve
32	176	54.3	190	10	Q8H732	Q8h732 kokia dryna
33	174	53.7	189	10	Q8H736	Q8h736 gossypium a
34	174	53.7	189	10	Q8H734	Q8h734 gossypium l
35	174	53.7	189	10	Q8H733	Q8h733 gossypium b
36	174	53.7	189	10	Q8GV62	Q8gv62 gossypium g
37	174	53.7	189	10	Q8GV61	Q8gv61 gossypium d
38	174	53.7	189	10	Q8GV60	Q8gv60 gossypium s
39	174	53.7	189	10	Q8GV58	Q8gv58 gossypium t
40	174	53.7	1129	10	Q8LK26	Q8lk26 mesotaenium
41	170	52.5	189	10	Q8GV59	Q8gv59 gossypium t
42	165	50.9	1043	10	Q9FNC3	Q9fnc3 arabidopsis
43	165	50.9	1055	10	Q8GZN8	Q8gzn8 arabidopsis
44	162	50.0	187	10	Q9M692	Q9m692 gossypioide
45	162	50.0	1075	10	Q9LLI9	Q9lli9 zea mays (m

ALIGNMENTS

RESULT 1
Q9LLJ2
ID Q9LLJ2 PRELIMINARY; PRT; 56 AA.
AC Q9LLJ2;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Cellulose synthase (Fragment).
GN CESA.
OS Populus tremuloides (Quaking aspen).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Malpighiales; Salicaceae; Populus.
OX NCBI_TaxID=3693;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Xylem;
RX MEDLINE=20344632; PubMed=10886769;
RA Wu L., Joshi C.P., Chiang V.L.;
RT "A xylem-specific cellulose synthase gene from aspen (Populus
RT tremuloides) is responsive to mechanical stress.";
RL Plant J. 22:495-502(2000).
DR EMBL; AF197911; AAF91155.1; -.
FT NON_TER 56 56
SQ SEQUENCE 56 AA; 6160 MW; 11BFFF8C37328402 CRC64;

Query Match 100.0%; Score 324; DB 10; Length 56;
Best Local Similarity 100.0%; Pred. No. 4e-34;
Matches 56; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MMESGAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEIKEGRKVCLRCGS 56
Db 1 MMESGAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEIKEGRKVCLRCGS 56

RESULT 2
O81368
ID O81368 PRELIMINARY; PRT; 978 AA.
AC O81368;

```
T   01-NOV-1998 (TrEMBLrel. 08, Created)
T   01-JAN-1999 (TrEMBLrel. 09, Last sequence update)
T   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
E   Secondary xylem cellulose synthase.
N   CELA.
S   Populus tremuloides (Quaking aspen).
C   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
C   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
C   eurosids I; Malpighiales; Salicaceae; Populus.
X   NCBI_TaxID=3693;
N   [1]
P   SEQUENCE FROM N.A.
C   TISSUE=Secondary xylem;
X   MEDLINE=20344632; PubMed=10886769;
A   Wu L., Joshi C.P., Chiang V.L.;
T   "A xylem-specific cellulose synthase gene from aspen (Populus
T   tremuloides) is responsive to mechanical stress.";
L   Plant J. 22:495-502(2000).
R   EMBL; AF072131; AAD03417.1; -.
R   InterPro; IPR005150; Cellulose_synt.
R   Pfam; PF03552; Cellulose_synt; 1.
Q   SEQUENCE   978 AA;  110279 MW;  C351B0E4975C6B8A CRC64;

 Query Match             98.5%;  Score 319;  DB 10;  Length 978;
 Best Local Similarity   98.2%;  Pred. No. 2.7e-32;
 Matches   55;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MMESGAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEIKEGRKVCLRCGS 56
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MMESGAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEINEGRKVCLRCGS 56


RESULT 3
P93155
ID   P93155      PRELIMINARY;      PRT;   974 AA.
AC   P93155;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Cellulose synthase.
GN   CELA1.
OS   Gossypium hirsutum (Upland cotton).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium.
OX   NCBI_TaxID=3635;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Acala SJ-2; TISSUE=Fiber;
RX   MEDLINE=97057296; PubMed=8901635;
RA   Pear J.R., Kawagoe Y., Schreckengost W.E., Delmer D.P., Stalker D.M.;
RT   "Higher plants contain homologs of the bacterial celA genes encoding
RT   the catalytic subunit of cellulose synthase.";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:12637-12642(1996).
DR   EMBL; U58283; AAB37766.1; -.
DR   InterPro; IPR005150; Cellulose_synt.
DR   Pfam; PF03552; Cellulose_synt; 1.
SQ   SEQUENCE   974 AA;  109703 MW;  2656CA2D05D100A7 CRC64;

 Query Match             81.5%;  Score 264;  DB 10;  Length 974;
 Best Local Similarity   75.0%;  Pred. No. 2.9e-25;
 Matches   42;  Conservative    8;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          1 MMESGAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEIKEGRKVCLRCGS 56
              ||||| |:|||||| || : ||| |||||||::|:||||||:::|||||| ||||||
Db          1 MMESGVPVCHTCGEHVGLNVNGEPFVACHECNFPICKSCFEYDLKEGRKACLRCGS 56


RESULT 4
Q8W1W0
ID   Q8W1W0      PRELIMINARY;      PRT;   974 AA.
AC   Q8W1W0;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Cellulose synthase A4
GN   CESA4.
OS   Gossypium hirsutum (Upland cotton).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium.
OX   NCBI_TaxID=3635;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21608363; PubMed=11743074;
RA   Kim H.J., Triplett B.A.;
RT   "Cotton Fiber Growth in Planta and in Vitro  Models for Plant Cell
RT   Elongation and Cell Wall Biogenesis.";
RL   Plant Physiol. 127:1361-1366(2001).
DR   EMBL; AF413210; AAL37[illegible].1; -.
DR   InterPro; IPR005150; Cellulose_synt.
DR   Pfam; PF03552; Cellulose_synt; 1.
SQ   SEQUENCE   974 AA;  [illegible] MW;  [illegible] CRC64;

 Query Match             81.5%;  Score 264;  DB 10;  Length 974;
 Best Local Similarity   75.0%;  Pred. No. 2.9e-25;
 Matches   42;  Conservative    8;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          1 MMESGAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEIKEGRKVCLRCGS 56
              ||||| |:|||||| || : ||| |||||||::|:||||||:::|||||| ||||||
Db          1 MMESGVPVCHTCGEHVGLNVNGEPFVACHECNFPICKSCFEYDLKEGRKACLRCGS 56


RESULT 5
Q9C5Z8
ID   Q9C5Z8      PRELIMINARY;      PRT;   985 AA
AC   Q9C5Z8;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Cellulose synthase catalytic subunit.
GN   IRX1.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21066939; PubMed=11148295;
RA   Taylor N.G., Laurie S., Turner S.R.;
RT   "Multiple Cellulose Synthase Catalytic Subunits Are Required for
RT   Cellulose Synthesis in Arabidopsis.";
RL   Plant Cell 12:2529-2540(2000).
DR   EMBL; AF267742; AAK08709.1; -.
DR   InterPro; IPR005150; Cellulose_synt.
DR   Pfam; PF03552; Cellulose_synt; 1.
SQ   SEQUENCE   985 AA;  111470 MW;  693C701B7B69AB25 CRC64;

 Query Match             76.8%;  Score 247;  DB 10;  Length 985;
 Best Local Similarity   67.9%;  Pred. No. 4.4e-23;
 Matches   38;  Conservative   12;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          1 MMESGAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEIKEGRKVCLRCGS 56
              |||| :|||:||.:::|  :||  ||| ||||:|||:| |:| .|||::|.|||:
Db          1 MMESRSPICNTCGFEIGAKSNGEPFVACHECSFPICKACLEYEFKEGRRICLRCGN 56


RESULT 6
Q9SN37
ID   Q9SN37      PRELIMINARY;      PRT;   958 AA.
AC   Q9SN37;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Cellulose synthase-like protein.
```

```
GN   F28A21.190 OR AT4G18780.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Bevan M., Mueller M.W., Muendlein A., Felber R., Bancroft I.,
RA   Mewes H.W., Mayer K.F.X., Lemcke K., Schueller C.;
RL   Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   EU Arabidopsis sequencing project;
RL   Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Mueller M.W., Muendlein A., Felber R., Mewes H.W., Lemcke K.,
RA   Mayer K.F.X.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   EU Arabidopsis sequencing project;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL035526; CAB37463.1; -.
DR   EMBL; AL161549; CAB78880.1; -.
DR   InterPro; IPR005150; Cellulose_synt
DR   Pfam; PF03552; Cellulose_synt; 1.
SQ   SEQUENCE   958 AA;  108520 MW;  B3836EF4BC697502 CRC64;

  Query Match             75.9%;  Score 246;  DB 10;  Length 958;
  Best Local Similarity   67.9%;  Pred. No. 5.7e-23;
  Matches   38;  Conservative   12;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          1 MMESGAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEIKEGRKVCLRCGS 56
              |||| :|||:||||::|  :|'| |||:||||:|:||:| |:| .|||::|||||:
Db          1 MMESRSPICNTCGEEIGVKSNGEFFVACHECSFPICKACLEYEFKEGRRICLRCGN 56


RESULT 7
Q8LPK5
ID   Q8LPK5      PRELIMINARY;      PRT;   985 AA.
AC   Q8LPK5;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Cellulose synthase-like protein.
GN   AT4G18780.
OS   Arabidopsis thaliana (Mouse-ear cress)
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Nguyen M., Karlin-Neumann G., Southwick A., Lam B., Miranda M.,
RA   Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA   Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA   Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA   Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
RA   Ecker J., Theologis A., Davis R.W.;
RL   Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY099636; AAM20487.1;
DR   InterPro; IPR005150; Cellulose_synt
DR   Pfam; PF03552; Cellulose_synt; 1.
SQ   SEQUENCE   985 AA;  111521 MW;  D299F357625768E0 CRC64;

  Query Match             75.9%;  Score 246;  DB 10;  Length 985;
  Best Local Similarity   67.9%;  Pred. No. 5.9e-23;
  Matches   38;  Conservative   12;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          1 MMESGAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEIKEGRKVCLRCGS 56
              |||| :|| :|||::|  :|:. |||'.|||:|:||:  |:| :|||::|||||:

Db          1 MMESRSPICNTCGEEIGVKSNGEFFVACHECSFPICKACLEYEFKEGRRICLRCGN 56


RESULT 8
Q9SYX4
ID   Q9SYX4      PRELIMINARY;      PRT;   205 AA.
AC   Q9SYX4;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Cellulose synthase A1 (Fragment).
GN   CELA1.
OS   Gossypium hirsutum (Upland cotton).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium.
OX   NCBI_TaxID=3635;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Cronn R.C., Small R.L., Wendel J.F.;
RT   "Duplicated genes evolve independently following polyploid formation
RT   in cotton.";
RL   Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF139443; AAD33796.1; -.
DR   InterPro; IPR005150; Cellulose_synt.
DR   Pfam; PF03552; Cellulose_synt; 1.
FT   NON_TER       1       1
FT   NON_TER     205     205
SQ   SEQUENCE   205 AA;  23118 MW;  0EEAF68904179C33 CRC64;

  Query Match             67.9%;  Score 220;  DB 10;  Length 205;
  Best Local Similarity   74.5%;  Pred. No. 2.8e-20;
  Matches   35;  Conservative    7;  Mismatches    5;  Indels    0;  Gaps    0;

Qy         10 HTCGEQVGHDANGELFVACHECSYPMCKSCFEFEIKEGRKVCLRCGS 56
              |||:| || : ||| |||| ||:::|:|||||||:::||||| ||||||
Db          1 HTCGEHVGLNVNGEPFVACHECNFPICKSCFEYDLKEGRKACLRCGS 47


RESULT 9
Q9XFC6
ID   Q9XFC6      PRELIMINARY;      PRT;   205 AA.
AC   Q9XFC6;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Cellulose synthase A1 (Fragment).
GN   CELA1.
OS   Gossypium robinsonii.
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium.
OX   NCBI_TaxID=3636;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Cronn R.C., Small R.L., Wendel J.F.;
RT   "Duplicated genes evolve independently following polyploid formation
RT   in cotton.";
RL   Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF139446; AAD33799.1; -.
DR   InterPro; IPR005150; Cellulose_synt.
DR   Pfam; PF03552; Cellulose_synt; 1.
FT   NON_TER       1       1
FT   NON_TER     205     205
SQ   SEQUENCE   205 AA;  23132 MW;  FF785E02BE179E69 CRC64;

  Query Match             67.9%;  Score 220;  DB 10;  Length 205;
  Best Local Similarity   74.5%;  Pred. No. 2.8e-20;
  Matches   35;  Conservative    7;  Mismatches    5;  Indels    0;  Gaps    0;

Qy         10 HTCGEQVGHDANGELFVACHECSYPMCKSCFEFEIKEGRKVCLRCGS 56
              |||:| || : ||| ||||||||::|:|||||||:::||||| ||||||
Db          1 HTCGEHVGLNVNGEPFVACHECNFPICKSCFEYDLKEGRKACLRCGS 47
```

```
ESULT 10
9XFC5
D   Q9XFC5      PRELIMINARY;      PRT;   205 AA.
C   Q9XFC5;
T   01-NOV-1999 (TrEMBLrel. 12, Created)
T   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
T   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
E   Cellulose synthase A1 (Fragment).
N   CELA1.
S   Gossypium raimondii.
C   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
C   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
C   eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium.
X   NCBI_TaxID=29730;
N   [1]
P   SEQUENCE FROM N.A.
A   Cronn R.C., Small R.L., Wendel J.F.;
T   "Duplicated genes evolve independently following polyploid formation
T   in cotton.";
L   Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases
R   EMBL; AF139444; AAD33797.1; -.
R   InterPro; IPR005150; Cellulose_synt.
R   Pfam; PF03552; Cellulose_synt; 1.
T   NON_TER       1      1
T   NON_TER     205    205
Q   SEQUENCE   205 AA;  23105 MW;  941ACA60F88A140C CRC64;

 Query Match             67.9%;  Score 220;  DB 10;  Length 205;
 Best Local Similarity   74.5%;  Pred. No. 2.8e-20;
 Matches   35;  Conservative    7;  Mismatches    5;  Indels    0;  Gaps    0;

Qy         10 HTCGEQVGHDANGELFVACHECSYPMCKSCFEFEIKEGRKVCLRCGS 56
              ||||| || : ||| |||||||::|:||||||:::||||| ||||||
Db          1 HTCGEHVGLNVNGEPFVACHECNFPICKSCFEYDLKEGRKACLRCGS 47


ESULT 11
9XFC4
D   Q9XFC4      PRELIMINARY;      PRT;   205 AA.
C   Q9XFC4;
T   01-NOV-1999 (TrEMBLrel. 12, Created)
T   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
T   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
E   Cellulose synthase A1 (Fragment).
N   CELA1.
S   Gossypium herbaceum.
C   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
C   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
C   eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium.
X   NCBI_TaxID=34274;
N   [1]
P   SEQUENCE FROM N.A.
A   Cronn R.C., Small R.L., Wendel J.F.;
T   "Duplicated genes evolve independently following polyploid formation
T   in cotton.";
L   Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
R   EMBL; AF139442; AAD33795.1; -.
R   InterPro; IPR005150; Cellulose_synt.
R   Pfam; PF03552; Cellulose_synt; 1.
T   NON_TER       1      1
T   NON_TER     205    205
Q   SEQUENCE   205 AA;  23118 MW;  0EEAF68904179C33 CRC64;

 Query Match             67.9%;  Score 220;  DB 10;  Length 205;
 Best Local Similarity   74.5%;  Pred. No. 2.8e-20;
 Matches   35;  Conservative    7;  Mismatches    5;  Indels    0;  Gaps    0;

Qy         10 HTCGEQVGHDANGELFVACHECSYPMCKSCFEFEIKEGRKVCLRCGS 56
              ||||| || : ||| |||||||::|:||||||:::||||| ||||||
Db          1 HTCGEHVGLNVNGEPFVACHECNFPICKSCFEYDLKEGRKACLRCGS 47
```

```
RESULT 12
Q9SYX5
ID   Q9SYX5      PRELIMINARY;      PRT;   205 AA.
AC   Q9SYX5;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Cellulose synthase A1 (Fragment).
GN   CELA1.
OS   Gossypium hirsutum (Upland cotton).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium.
OX   NCBI_TaxID=3635;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Cronn R.C., Small R.L., Wendel J.F.;
RT   "Duplicated genes evolve independently following polyploid formation
RT   in cotton.";
RL   Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases
DR   EMBL; AF139445; AAD33796.1; -.
DR   InterPro; IPR005150; Cellulose_synt.
DR   Pfam; PF03552; Cellulose_synt; 1.
FT   NON_TER       1      1
FT   NON_TER     205    205
SQ   SEQUENCE   205 AA;  23079 MW;  603CA29F9F06D656 CRC64;

  Query Match             66.1%;  Score 211;  DB 10;  Length 205;
  Best Local Similarity   71.2%;  Pred. No. 3.9e-19;
  Matches   33;  Conservative    9;  Mismatches    5;  Indels    0;  Gaps    0;

Qy         10 HTCGEQVGHDANGELFVACHECSYPMCKSCFEFEIKEGRKVCLRCGS 56
              || || || : : | |::|||::|:|:.||::: |||| :|||:
Db          1 HTCGEHVGLNVSGEPFVACHECNFPICESCFEYDLKEGRKACLRCGS 47


RESULT 13
Q9AXK0
ID   Q9AXK0      PRELIMINARY;      PRT;   978 AA.
AC   Q9AXK0;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Cellulose synthase CesA 1.
OS   Zinnia elegans.
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC   Asteridae; campanulids; Asterales; Asteraceae; Asteroideae;
OC   Heliantheae; Zinnia.
OX   NCBI_TaxID=34245;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21438255; PubMed=11554477;
RA   Haigler C.H., Ivanova-Datcheva M., Hogan P.S., Salnikov V.V.,
RA   Hwang S., Martin K., Delmer D.P.;
RT   "Carbon partitioning to cellulose synthesis.";
RL   Plant Mol. Biol. 47:29-51(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Hwang S., Haigler C.H.;
RL   Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF323039; AAK11588.2; -.
DR   InterPro; IPR005150; Cellulose_synt.
DR   Pfam; PF03552; Cellulose_synt; 1.
SQ   SEQUENCE   978 AA;  110353 MW;  399D76EF33427B1D CRC64;

  Query Match             59.7%;  Score 193.5;  DB 10;  Length 978;
  Best Local Similarity   55.2%;  Pred. No. 3e-16;
  Matches   32;  Conservative   10;  Mismatches   13;  Indels    3;  Gaps    2;

Qy          1 MMESGAPICHTCGEQVGH--DANGELFVACHECSYPMCKSCFEFEIKEGRKVCLRCGS 56
              ||    |||:|||| :|   | || :|| ||||::|:|||| |:|| ||   |::|||
```

```
Db          1 MMHPETPICNTCGEHIGFGPDGNG-VFVGCHECNFPICKSCLEYEINEGHTACIQCGT 57


RESULT 14
Q93WX2
ID   Q93WX2      PRELIMINARY;      PRT;   197 AA.
AC   Q93WX2;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Cellulose synthase A1 (Fragment).
GN   CESA1.
OS   Gossypium somalense.
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium.
OX   NCBI_TaxID=47635;
RN   [1]
RP   SEQUENCE FROM N.A
RC   TISSUE=Fiber;
RA   Cronn R.C., Small R.L., Haselkorn T., Wendel J.F.;
RT   "Rapid diversification of the cotton genus (Gossypium:Malvaceae)
RT   revealed by analysis of sixteen nuclear and chloroplast genes.";
RL   Am. J. Bot. 0:0-0(2001).
DR   EMBL; AF419971; AAL17601.1; -.
DR   InterPro; IPR005150; Cellulose_synt
DR   Pfam; PF03552; Cellulose_synt; 1.
FT   NON_TER       1      1
FT   NON_TER     197    197
SQ   SEQUENCE   197 AA;  22094 MW;  FFAB668DFB8CF612 CRC64;

  Query Match             59.3%;  Score 192;  DB 10;  Length 197;
  Best Local Similarity   72.1%;  Pred. No. 1e-16;
  Matches   31;  Conservative    7;  Mismatches    5;  Indels    0;  Gaps    0;

Qy         14 EQVGHDANGELFVACHECSYPMCKSCFEFEIKEGRKVCLRCGS 56
              | ||: : ||| |||||||:::|||:|:::|||| ||||'
Db          3 EHVGLNVNGEPFVACHECNFPICKSCFEYDLKEGRKACLRCGS 45


RESULT 15
Q8GV66
ID   Q8GV66      PRELIMINARY;      PRT;   197 AA.
AC   Q8GV66;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Cellulose synthase A1 (Fragment).
GN   CESA1.
OS   Gossypium davidsonii.
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium
OX   NCBI_TaxID=34287;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Cronn R., Small R., Haselkorn T., Wendel J.,
RT   "Cryptic repeated genomic recombination during speciation in
RT   Gossypium.";
RL   Submitted (JUN-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY125066; AAM98391.1; -.
FT   NON_TER       1      1
FT   NON_TER     197    197
SQ   SEQUENCE   197 AA;  22084 MW;  FE1ACE35AB5A73B1 CRC64;

  Query Match             59.3%;  Score 192;  DB 10;  Length 197;
  Best Local Similarity   72.1%;  Pred. No. 1e-16;
  Matches   31;  Conservative    7;  Mismatches    5;  Indels    0;  Gaps    0;

Qy         14 EQVGHDANGELFVACHECSYPMCKSCFEFEIKEGRKVCLRCGS 56
              | ||: : | | |||||||::|:|||:|!:::| ||| :||| :
Db          3 EHVGLNVNGEPFVACHECNFPICKSCFEYDLKEGRKACLRCGS 45
```

Search completed: October 29, 2003, 18:30:51
Job time : 4.89942 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 29, 2003, 18:25:23 ; Search time 1.19149 Seconds
(without alignments)
2210.255 Million cell updates/sec

Title: US-09-980-043-4
Perfect score: 324
Sequence: 1 MMESGAPICHTCGEQVGHLA.....KSCFEFEIKEGRKVCLRCGS 56

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026725 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No	Score	% Query Match	Length	DB	ID	Description
1	80	24.7	485	1	RN14_MOUSE	Q9ji90 mus musculu
2	74	22.8	560	1	SET4_YEAST	P42948 saccharomyc
3	72	22.2	474	1	RN14_HUMAN	Q9ubs8 homo sapien
4	71.5	22.1	1958	1	UBR1_SCHPO	O60152 schizosacch
5	68.5	21.1	570	1	GRAU_DROME	Q9u405 drosophila
6	68	21.0	172	1	CNBP_CHICK	O42395 gallus gall
7	67	20.7	367	1	ZZ11_HUMAN	Q13398 homo sapien
8	67	20.7	519	1	ZN24_HUMAN	P49682 homo sapien
9	67	20.7	681	1	NPA4_MOUSE	P49708 mus musculu
10	67	20.7	684	1	NPA4_RAT	P49709 rattus norv
11	67	20.7	3866	1	HRX_MOUSE	P55200 mus musculu
12	67	20.7	3969	1	HRX_HUMAN	Q03164 homo sapien
13	66.5	20.5	309	1	MAT1_HUMAN	P51948 homo sapien
14	66.5	20.5	309	1	MAT1_MOUSE	P51949 mus musculu
15	66	20.4	647	1	LIK1_HUMAN	P53667 homo sapien
16	65.5	20.2	325	1	PINC_HUMAN	P48059 homo sapien
17	65.5	20.2	329	1	PIN2_CAEEL	Q19157 caenorhabdi
18	65	20.1	647	1	LIK1_MOUSE	P53668 mus musculu
19	65	20.1	704	1	RP3A_BOVIN	Q06846 bos taurus
20	65	20.1	748	1	AD10_HUMAN	O14672 homo sapien
21	64.5	19.9	309	1	MAT1_XENLA	P51951 xenopus lae
22	64	19.8	279	1	SLI3_MOUSE	O70433 mus musculu
23	64	19.8	279	1	SLI3_RAT	O35115 rattus norv
24	64	19.8	340	1	IS2B_ONCTS	P50212 oncorhynchu
25	64	19.8	359	1	ISL2_BRARE	P53406 brachydanio
26	64	19.8	694	1	RP3A_HUMAN	Q9y2j0 homo sapien
27	63.5	19.6	736	1	TOP1_HELPY	P56991 helicobacte
28	63.5	19.6	1042	1	LI149_CAEEL	Q20318 caenorhabdi
29	63	19.4	358	1	IS2A_ONCTS	P52408 oncorhynchu
30	63	19.4	363	1	ISL3_ONCTS	P52409 oncorhynchu
31	63	19.4	376	1	VE1_HPV28	P50771 human papil
32	63	19.4	383	1	VE1_HPV01	P06778 human papil
33	63	19.4	738	1	PRI1_HUMAN	P49754 homo sapien
34	62.5	19.3	179	1	BYR3_SCHPO	P36627 schizosacch
35	62.5	19.3	736	1	TOP1_HELPJ	Q9zmv7 helicobacte
36	62	19.1	423	1	TES_MOUSE	P47226 mus musculu
37	62	19.1	482	1	ARI2_CAEEL	Q22431 caenorhabdi
38	62	19.1	544	1	AD10_RAT	Q10743 rattus norv
39	62	19.1	545	1	Z307_HUMAN	Q969j2 homo sapien
40	62	19.1	749	1	AD10_MOUSE	O35598 mus musculu
41	62	19.1	1696	1	PCK5_BRACL	Q9nj15 branchiosto
42	61.5	19.0	170	1	CNBP_MOUSE	P53996 mus musculu
43	61.5	19.0	177	1	CNBP_HUMAN	P20694 homo sapien
44	61.5	19.0	1959	1	AGRI_RAT	P25304 rattus norv
45	61	18.8	411	1	ASLB_ECOLI	P25550 escherichia

ALIGNMENTS

RESULT 1
RN14_MOUSE
ID RN14_MOUSE STANDARD; PRT; 485 AA.
AC Q9JI90; Q9D0L2; Q9D6N2; Q9D6Z8; Q9JI89;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE RING finger protein 14 (Androgen receptor-associated protein 54)
DE (Triad2 protein).
GN RNF14 OR ARA54 OR TRIAD2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RA Van der Reijden B.A., Jansen J.H.;
RT "Identification of a novel TRIAD protein, TRIAD2.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC STRAIN=C57BL/6J; TISSUE=Embryo, and Tongue;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
CC -!- FUNCTION: MIGHT ACT AS AN E3 UBIQUITIN-PROTEIN LIGASE WHICH
CC ACCEPTS UBIQUITIN FROM SPECIFIC E2 UBIQUITIN-CONJUGATING ENZYMES
CC AND THEN TRANSFERS IT TO SUBSTRATES, WHICH COULD BE NUCLEAR
CC PROTEINS. COULD PLAY A ROLE AS A COACTIVATOR FOR ANDROGEN AND, TO
CC A LESSER EXTENT, PROGESTERONE-DEPENDENT TRANSCRIPTION (BY
CC SIMILARITY).
CC -!- SUBUNIT: INTERACTS WITH THE UBIQUITIN-CONJUGATING ENZYMES UBE2E1
CC AND UBE2E2 AND, IN THE PRESENCE OF TESTOSTERONE, WITH THE ANDROGEN
CC RECEPTOR (AR) (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Nuclear and cytoplasmic (By similarity).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;

```
C         IsoId=Q9JI90-1; Sequence=Displayed;
C       Name=2;
C         IsoId=Q9JI90-2; Sequence=VSP_005751, VSP_005752;
C    -!- DOMAIN: THE N-TERMINAL DESTRUCTION BOX (D-BOX) COULD ACT AS A
C       RECOGNITION SIGNAL FOR DEGRADATION VIA THE UBIQUITIN-PROTEASOME
C       PATHWAY (BY SIMILARITY).
C    -!- DOMAIN: THE RING-TYPE ZINC FINGER IS ESSENTIAL FOR THE INTERACTION
C       WITH UBE2E2 (BY SIMILARITY).
C    -!- PTM: RING-TYPE ZINC FINGER-DEPENDENT AND UBE2E2-DEPENDENT
C       AUTOUBIQUITINATION (BY SIMILARITY).
C    -!- SIMILARITY: Contains 1 IBR-type zinc finger.
C    -!- SIMILARITY: Contains 1 RING-type zinc finger.
C    -!- SIMILARITY: Contains 1 RWD domain.
C    -----------------------------------------------------------------------
C    This SWISS-PROT entry is copyright. It is produced through a collaboration
C    between the Swiss Institute of Bioinformatics and the EMBL outstation -
C    the European Bioinformatics Institute. There are no restrictions on its
C    use by non-profit institutions as long as its content is in no way
C    modified and this statement is not removed. Usage by and for commercial
C    entities requires a license agreement (See http://www.isb-sib.ch/announce/
C    or send an email to license@isb-sib.ch)
C    -----------------------------------------------------------------------
R    EMBL; AF249667; AAF74266.1; -.
R    EMBL; AF249668; AAF74267.1; -.
R    EMBL; AK011316; BAB27541.1; -.
R    EMBL; AK009783; BAB26503.1; -.
R    EMBL; AK010162; BAB26741.1; -.
R    MGD; MGI:1929668; Rnf14.
R    GO; GO:0005737; C:cytoplasm; ISS.
R    GO; GO:0005634; C:nucleus; ISS.
R    GO; GO:0005515; F:protein binding activity; ISS.
R    GO; GO:0003713; F:transcription co-activator activity; ISS.
R    GO; GO:0019787; F:ubiquitin-like-protein ligase activity; ISS.
R    GO; GO:0006357; P:regulation of transcription from Pol II pro. . .; ISS.
R    GO; GO:0007165; P:signal transduction; ISS.
R    InterPro; IPR006575; RWD.
R    InterPro; IPR002867; Znf_C6HC.
R    InterPro; IPR001841; Znf_ring.
R    Pfam; PF01485; IBR; 1.
R    Pfam; PF00097; zf-C3HC4; 1.
R    SMART; SM00647; IBR; 2.
R    SMART; SM00591; RWD; 1.
R    PROSITE; PS50908; RWD; 1.
DR   PROSITE; PS00518; ZF_RING_1; 1.
DR   PROSITE; PS50089; ZF_RING_2; 1.
KW   Ligase; Ubl conjugation pathway; Transcription regulation;
KW   Zinc-finger; Nuclear protein; Alternative splicing.
FT   DOMAIN       11    137       RWD.
FT   ZN_FING     221    267       RING-TYPE.
FT   ZN_FING     290    351       IBR-TYPE.
FT   DOMAIN       37     45       D-BOX.
FT   DOMAIN      173    176       POLY-ALA.
FT   DOMAIN      470    485       ASP/GLU-RICH (ACIDIC).
FT   VARSPLIC     52     79       GNSNESLQNSGFEYTICFLPPLVLNFEL -> ALCSLQAPR
FT                                QPVGRTPRQSGAVCLDAVS (in isoform 2).
FT                                /FTId=VSP_005751.
FT   VARSPLIC     80    485       Missing (in isoform 2).
FT                                /FTId=VSP_005752.
FT   CONFLICT     22     22       V -> A (IN REF. 2).
FT   CONFLICT     62     62       G -> R (IN REF. 2; BAB26503).
FT   CONFLICT    103    106       LSAL -> GTLV (IN REF. 2; BAB26503/
FT                                BAB26741).
SQ   SEQUENCE   485 AA;  54954 MW;  C9430796EB47112B CRC64;

  Query Match             24.7%;  Score 80;  DB 1;  Length 485;
  Best Local Similarity   36.5%;  Pred. No. 0.048;
  Matches   19;  Conservative    8;  Mismatches   15;  Indels   10;  Gaps    3;

Qy          8 ICHTC-GEQVGHDANGELFVACHECSYPMCKSC----FEFEIKEGRKVCLRC 54
              :|  |  |::| |    |     || :  ||:|    || :||:|:  || |
Db        220 LCSICFCEKLGSDCMYFL-----ECKHVYCKACLKDYFEIQIKDGQVKCLNC 266

RESULT 2
SET4_YEAST
ID   SET4_YEAST     STANDARD;      PRT;   560 AA.
AC   P42948;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   SET domain protein 4.
GN   SET4 OR YCL105W OR C0481.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Asco[illegible]; Saccharomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / FY1679;
RX   MEDLINE=96091136; PubMed=8603841;
RA   Rasmussen S.W.;
RT   "A [illegible] kb region [illegible] codes the SME1, MRP2, GSH1
RT   and CDC8 genes, [illegible] encoding tRNA genes, and [illegible]
RT   to the DAL80 [illegible] DNA.";
RL   Yeast 11:873-884(1995).
CC   !- FUNCTION: Putative chromatin regulator.
CC   -!- SIMILARITY: Belongs to the SET3 family.
CC   -!- SIMILARITY: Contains 1 PHD-type zinc finger.
CC   -!- SIMILARITY: Contains 1 SET domain.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; X85021; CAA59389.1; -.
DR   EMBL; Z49380; CAA89403.1; -.
DR   PIR; S53382; S53382.
DR   SGD; S0003641; SET4.
DR   InterPro; IPR001214; SET.
DR   InterPro; IPR001965; Znf_PHD.
DR   Pfam; PF00628; PHD; 1.
DR   Pfam; PF00856; SET; 1.
DR   SMART; SM00249; PHD; 1.
DR   SMART; SM00317; SET; 1.
DR   PROSITE; PS50280; SET; 1.
DR   PROSITE; PS01359; ZF_PHD_1; 1.
DR   PROSITE; PS50016; ZF_PHD_2; 1.
KW   Chromatin regulator; Zinc-finger.
FT   ZN_FING     160    210       PHD-TYPE.
FT   DOMAIN      345    479       SET.
SQ   SEQUENCE   560 AA;  63805 MW;  9A7BAECF97418AFA CRC64;

  Query Match             2.8%;  Score 74;  DB 1;  Length 560;
  Best Local Similarity   35.4%;  Pred. No. 0.27;
  Matches   17;  Conservative    7;  Mismatches   18;  Indels    6;  Gaps    2;

Qy         12 CGEQVGHDANGELFVACHECSYPMCKSCFEFEIKEGRK---VCLRCGS 56
              ||     |:  |||. |::|     ' |: :  :   :   || || :
Db        165 CGSF---DSKDELF QCNKCKTWQHKLCYAFFKSDFEIKRDFVCKRCDS 209

RESULT 3
RN14_HUMAN
ID   RN14_HUMAN     STANDARD;      PRT;   474 AA.
AC   Q9UBS8; O94793;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   RING finger protein 14 (Androgen receptor associated protein 54)
DE   (Triad2 protein) (HFB30) (HRIHFB2038).
GN   RNF14 OR ARA54.
OS   Homo sapiens (Human).
```

```
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Fetal brain;
RX   MEDLINE=99255429; PubMed=10320776;
RA   Ueki N., Seki N., Yano K., Masuho Y., Saito T., Muramatsu M.-A.;
RT   "Isolation and characterization of a novel human gene (RFB30) which
RT   encodes a protein with a RING finger motif.";
RL   Biochim. Biophys. Acta 1445:232-236(1999).
RN   [2]
RP   SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC   TISSUE=Prostate;
RX   MEDLINE=99185075; PubMed=10085091;
RA   Kang H.-Y., Yeh S., Fujimoto N., Chang C.;
RT   "Cloning and characterization of human prostate coactivator ARA54, a
RT   novel protein that associates with the androgen receptor.";
RL   J. Biol. Chem. 274:8570-8576(1999).
RN   [3]
RP   SEQUENCE OF 286-474 FROM N.A., AND SUBCELLULAR LOCATION.
RC   TISSUE=Fetal brain;
RX   MEDLINE=99068504; PubMed=9853615;
RA   Ueki N., Oda T., Kondo M., Yano K., Noguchi T., Muramatsu M.-A.;
RT   "Selection system for genes encoding nuclear-targeted proteins.";
RL   Nat. Biotechnol. 16:1338-1342(1998).
RN   [4]
RP   MUTAGENESIS OF CYS-220.
RX   MEDLINE=21223029; PubMed=11322894;
RA   Ito K., Adachi S., Iwakami R., Yasuda H., Muto Y., Seki N., Okano Y.;
RT   "N-terminally extended human ubiquitin-conjugating enzymes (E2s)
RT   mediate the ubiquitination of RING-finger proteins, ARA54 and RNF8.";
RL   Eur. J. Biochem. 268:2725-2732(2001).
CC   -!- FUNCTION: MIGHT ACT AS AN E3 UBIQUITIN-PROTEIN LIGASE WHICH
CC       ACCEPTS UBIQUITIN FROM SPECIFIC E2 UBIQUITIN-CONJUGATING ENZYMES
CC       AND THEN TRANSFERS IT TO SUBSTRATES, WHICH COULD BE NUCLEAR
CC       PROTEINS. COULD PLAY A ROLE AS A COACTIVATOR FOR ANDROGEN-AND, TO
CC       A LESSER EXTENT, PROGESTERONE DEPENDENT TRANSCRIPTION.
CC   -!- SUBUNIT: INTERACTS WITH THE UBIQUITIN-CONJUGATING ENZYMES UBE2E1
CC       AND UBE2E2 AND, IN THE PRESENCE OF TESTOSTERONE, WITH THE ANDROGEN
CC       RECEPTOR (AR).
CC   -!- SUBCELLULAR LOCATION: Nuclear and cytoplasmic.
CC   -!- TISSUE SPECIFICITY: Widely expressed.
CC   -!- DOMAIN: THE N-TERMINAL DESTRUCTION BOX (D-BOX) ACTS AS A
CC       RECOGNITION SIGNAL FOR DEGRADATION VIA THE UBIQUITIN-PROTEASOME
CC       PATHWAY (BY SIMILARITY).
CC   -!- DOMAIN: THE RING-TYPE ZINC FINGER IS ESSENTIAL FOR THE INTERACTION
CC       WITH UBE2E2.
CC   -!- PTM: RING-TYPE ZINC FINGER IS ESSENTIAL FOR THE E3
CC       AUTOUBIQUITINATION.
CC   -!- SIMILARITY: Contains 1 IBR type zinc finger.
CC   -!- SIMILARITY: Contains 1 RING type zinc finger.
CC   -!- SIMILARITY: Contains 1 RWD domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB022663; BAA78677.1; -.
DR   EMBL; AF060544; AAD21842.1; -.
DR   EMBL; AB015333; BAA34792.1; -.
DR   Genew; HGNC:10058; RNF14.
DR   MIM; 605675; -.
DR   GO; GO:0005737; C:cytoplasm; IDA.
DR   GO; GO:0005634; C:nucleus; IDA.
DR   GO; GO:0005515; F:protein binding; IPI.
DR   GO; GO:0003713; F:transcription coactivator activity; TAS.
DR   GO; GO:0019787; F:ubiquitin-like-protein ligase activity; IDA.
DR   GO; GO:0030521; P:androgen receptor signaling pathway; NAS.
DR   GO; GO:0006357; P:regulation of transcription from Pol II pro. . .; TAS.
DR   GO; GO:0007165; P:signal transduction; TAS.
DR   GO; GO:0006512; P:ubiquitin cycle; IEP.
DR   InterPro; IPR006575; RWD.
DR   InterPro; IPR002867; Znf_C6HC.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF01485; IBR; 1.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00647; IBR; 2.
DR   SMART; SM00591; RWD; 1.
DR   PROSITE; PS50908; RWD; 1.
DR   PROSITE; PS00518; ZF_RING_1; 1.
DR   PROSITE; PS50089; ZF_RING_2; 1.
KW   Ligase; Ubl conjugation pathway; Transcription regulation;
KW   Zinc-finger; Nuclear protein; Coiled coil.
FT   DOMAIN       11    137       RWD.
FT   ZN_FING     220    266       RING-TYPE.
FT   ZN_FING     289    350       IBR-TYPE.
FT   DOMAIN      351    395       COILED COIL (POTENTIAL).
FT   DOMAIN       37     45       D-BOX.
FT   DOMAIN      361    474       INTERACTION WITH ANDROGEN RECEPTOR.
FT   MUTAGEN     220    220       C->S: LOSS OF INTERACTION WITH UBE2E2 AND
FT                                OF AUTOUBIQUITINATION.
SQ   SEQUENCE   474 AA;  53837 MW;  529E3F5AF38A5DAD CRC64;

  Query Match             22.2%;  Score 72;  DB 1;  Length 474;
  Best Local Similarity   32.7%;  Pred. No. 0.39;
  Matches   17;  Conservative   10;  Mismatches   15;  Indels   10;  Gaps    3;

Qy          8 ICHTC-GEQVGHDANGELFVACHECSYPMCKSC----FEFEIKEGRKVCLRC 54
              :|  |  |::| :    |     || :  ||:|    || :|::|:  || |
Db        219 LCSICFCEKLGSECMYFL-----ECRHVYCKACLKDYFEIQIRDGQVQCLNC 265


RESULT 4
UBR1_SCHPO
ID   UBR1_SCHPO     STANDARD;      PRT;  1958 AA.
AC   O60152;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   N-end-recognizing protein (Ubiquitin-protein ligase E3 component) (N-
DE   recognin).
GN   UBR1 OR SPBC19C7.02 OR SPBC32F12.14.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21559218; PubMed=11702950;
RA   Kitamura K., Katayama S., Dhut S., Sato M., Watanabe Y., Yamamoto M.,
RA   Toda T.;
RT   "Phosphorylation of Mei2 and Ste11 by Pat1 kinase inhibits sexual
RT   differentiation via ubiquitin proteolysis and 14-3-3 protein in
RT   fission yeast.";
RL   Dev. Cell 1:389-399(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RX   MEDLINE=21848401; PubMed=11859360;
RA   Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA   Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA   Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA   Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA   Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA   Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA   James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA   Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA   Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,
RA   Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA   Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
```

```
A   Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
A   Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
A   Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
A   Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
A   Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
A   Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
A   Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
A   Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
A   Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
A   Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
A   Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
A   Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
A   Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
T   "The genome sequence of Schizosaccharomyces pombe.";
L   Nature 415:871-880(2002).
'C  -!- FUNCTION: RECOGNITION COMPONENT OF THE N-END RULE PATHWAY. BINDS
C       TO PROTEINS BEARING AMINO-TERMINAL RESIDUES THAT ARE DESTABILIZING
C       ACCORDING TO THE N-END RULE, BUT DOES NOT BIND TO OTHERWISE
C       IDENTICAL PROTEINS BEARING STABILIZING AMINO-TERMINAL RESIDUES (BY
C       SIMILARITY).
C   ----------------------------------------------------------------------- ----
C   This SWISS-PROT entry is copyright. It is produced through a collaboration
C   between the Swiss Institute of Bioinformatics and the EMBL outstation -
C   the European Bioinformatics Institute. There are no restrictions on its
C   use by non-profit institutions as long as its content is in no way
C   modified and this statement is not removed. Usage by and for commercial
C   entities requires a license agreement (See http://www.isb-sib.ch/announce/
C   or send an email to license@isb-sib.ch).
C   ----------------------------------------------------------------------------
R   EMBL; AB079542; BAB84667.1; -.
)R  EMBL; AL023859; CAA19570.1; -.
)R  EMBL; AL023796; CAA19375.1; -.
)R  PIR; T39808; T39808.
)R  GeneDB_SPombe; SPBC19C7.02; -.
)R  InterPro; IPR003126; Znf_Nrecognin.
)R  Pfam; PF02207; zf-UBR1; 1.
)R  SMART; SM00396; ZnF_UBR1; 1.
GW  Ligase; Ubl conjugation pathway.
SQ  SEQUENCE   1958 AA;  225756 MW;  65ABBB2ADC5911B5 CRC64;

  Query Match             22.1%;  Score 71.5;  DB 1;  Length 1958;
  Best Local Similarity   29.5%;  Pred. No. 1.7;
  Matches   18;  Conservative    5;  Mismatches   23;  Indels   15;  Gaps    3;

Qy          11 TCGEQVGH-DANGELFVACHECSY----PMCKSCF----------EFEIKEGRKVCLRCG 55
               ||  : ||    ||:|  |  ||       :|  ||          | :  |   |  ||
Db          91 TCESKCGHIFRKGEVFYRCKTCSVDSNSALCVKCFRATSHKDHETSFTVSAGSGGCCDCG 150

Qy          56 S 56
               :
Db         151 N 151


RESULT 5
GRAU_DROME
ID   GRAU_DROME     STANDARD;      PRT;   570 AA.
AC   Q9U405; Q9W2N7;
DT   15-SEP-2003 (Rel. 42, Created)
DT   15-SEP-2003 (Rel. 42, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Transcription factor grauzone.
GN   GRAU OR CG3282.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A., FUNCTION, TISSUE SPECIFICITY, AND MUTANTS QQ36 AND
RP   QE70.
RX   MEDLINE=20148536; PubMed=10683177;
RA   Chen B., Harms E., Chu T., Henrion G., Strickland S.;
RT   "Completion of meiosis in Drosophila oocytes requires transcriptional
RT   control by grauzone, a new zinc finger protein.";
RL   Development 127:1243 [illegible](2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [3]
RP   REVISIONS.
RC   STRAIN=Berkeley;
RX   MEDLINE=22426069; PubMed=12537572;
RA   Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA   Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA   Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA   Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,
RA   Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA   Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA   Lewis S.E.;
RT   "Annotation of the Drosophila melanogaster euchromatic genome: a
RT   systematic review.";
RL   Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley; TISSUE=Head;
RX   MEDLINE=22426066; PubMed=12537569;
RA   Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
RA   George R.A., Guarin H., Kronmiller B., Pacleb J.M., Park S., Wan K.H.,
RA   Rubin G.M., Celniker S.E.;
RT   "A Drosophila full-length cDNA resource.";
RL   Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).
RN   [5]
RP   FUNCTION, AND DNA-BINDING.
RX   MEDLINE=20384592; PubMed=10924478;
RA   Harms E., Chu T., Henrion G., Strickland S.;
RT   "The only function of Grauzone required for Drosophila oocyte meiosis
RT   is transcriptional activation of the cortex gene.";
```

```
RL   Genetics 155:1831-1839(2000).
CC   -!- FUNCTION: Transcription factor involved in female meiosis.
CC       Specifically required to bind to the promoter region of the Cortex
CC       gene and activate its expression in ovaries.
CC    -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- TISSUE SPECIFICITY: Present throughout embryogenesis, in larvae,
CC       in ovaries and other tissues of adult females, and in adult males.
CC       In ovaries, it is expressed in both follicle cells and nurse
CC       cells. Expressed from stage 9 in the germarium. Weakly or not
CC       expressed in the oocyte nucleus.
CC   -!- DEVELOPMENTAL STAGE: Expressed both maternally and zygotically.
CC   -!- SIMILARITY: Contains 7 C2H2-type zinc fingers.
CC   ---------- ---------- ---- ---- ----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------- ---------- ---- ---- ----
DR   EMBL; AF208016; AAF25356.1; -.
DR   EMBL; AE003452; AAF46653.2; -.
DR   EMBL; AY047537; AAK77269.1; -.
DR   FlyBase; FBgn0001133; grau.
DR   InterPro; IPR007087; Znf_C2H2.
DR   Pfam; PF00096; zf-C2H2; 8.
DR   SMART; SM00355; ZnF_C2H2; 8.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 7.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 7.
KW   Meiosis; DNA-binding; Transcription regulation; Activator;
KW   Metal-binding; Zinc; Zinc-finger; Repeat.
FT   ZN_FING      296     319       C2H2-TYPE 1.
FT   ZN_FING      327     349       C2H2-TYPE 2.
FT   ZN_FING      387     409       C2H2-TYPE 3.
FT   ZN_FING      417     441       C2H2-TYPE 4.
FT   ZN_FING      448     471       C2H2-TYPE 5.
FT   ZN_FING      476     498       C2H2-TYPE 6.
FT   ZN_FING      504     527       C2H2-TYPE 7.
FT   DOMAIN       293     570       DNA BINDING.
FT   DOMAIN       383     399       BIPARTITE NUCLEAR LOCALIZATION SIGNAL 1
FT                                  (POTENTIAL).
FT   DOMAIN       523     539       BIPARTITE NUCLEAR LOCALIZATION SIGNAL 2
FT                                  (POTENTIAL).
FT   DOMAIN       146     172       ASP-GLU-RICH (ACIDIC).
FT   MUTAGEN      298     298       C->Y: IN QE36; DEFECT IN FEMALE MEIOSIS.
FT                                  ABOLISHES DNA-BINDING ACTIVITY.
FT   MUTAGEN      493     493       E->K: IN QE79; DEFECT IN FEMALE MEIOSIS.
FT                                  ABOLISHES DNA-BINDING ACTIVITY.
SQ   SEQUENCE   570 AA;  65967 MW;  38B550B964C1EA7C CRC64;

  Query Match             21.1%;  Score 68.5;  DB 1;  Length 570;
  Best Local Similarity   30.4%;  Pred. No. 1.2;
  Matches   21;  Conservative    9;  Mismatches   20;  Indels   19;  Gaps    4;

Qy          6 APICHTCGEQVGHDANGELFVACH--------ECSYPMCKSCFEFE-IK--------EG 47
              ! ||| |;:::  :  |  :  |         :|  |  :| : |  :|         :|
Db        386 ANICHVCGKKIKDKAVFEKHVRLHFEESGPKIKCPRPDCRSALKDEDNLKQHLRRHNDEG 445

Qy         48 RK-VCLRCG 55
              :  :|  |:
Db        446 KLPICSECG 454


RESULT 6
CNBP_CHICK
ID   CNBP_CHICK     STANDARD;      PRT;   172 AA.
AC   O42395;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Cellular nucleic acid binding protein (CNBP).
GN   ZNF9 OR CNBP.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98129058; PubMed=9467877;
RA   van Heumen W.R.A., Claxton C., Pickles J.O.;
RT   "Sequence and tissue distribution of chicken cellular nucleic acid
RT   binding protein cDNA.";
RL   Comp. Biochem. Physiol. 118B:659-665(1997).
CC   -!- FUNCTION: SINGLE STRANDED DNA-BINDING PROTEIN, WITH SPECIFICITY TO
CC       THE STEROL REGULATORY ELEMENT (SRE). CNBP IS INVOLVED IN STEROL-
CC       MEDIATED REPRESSION (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC, ALSO PRESENT IN ENDOPLASMIC
CC       RETICULUM (BY SIMILARITY).
CC   -!- SIMILARITY: TO S.POMBE BYR3 AND TO RETROVIRAL NUCLEIC ACID BINDING
CC       PROTEINS (NBP).
CC   -!- SIMILARITY: Contains 7 CCHC-type zinc fingers.
CC   ---------- ---------- ---- ---- ----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------- ---------- ---- ---- ----
DR   EMBL; AF004942; AAB62243.1; -.
DR   InterPro; IPR001878; Znf_CCHC.
DR   Pfam; PF00098; zf-CCHC; 7.
DR   PRINTS; PR00939; C2HCZNFINGER.
DR   SMART; SM00343; ZnF_C2HC; 7.
DR   PROSITE; PS50158; ZF_CCHC; 7.
KW   Zinc-finger; DNA-binding; Transcription regulation; Repressor; Repeat.
FT   ZN_FING       4      21       CCHC-TYPE 1.
FT   ZN_FING      46      63       CCHC-TYPE 2.
FT   ZN_FING      67      84       CCHC-TYPE 3.
FT   ZN_FING      91     108       CCHC-TYPE 4.
FT   ZN_FING     112     129       CCHC-TYPE 5.
FT   ZN_FING     130     147       CCHC-TYPE 6.
FT   ZN_FING     151     168       CCHC-TYPE 7.
SQ   SEQUENCE   172 AA;  19043 MW;  123F4E248A980390 CRC64;

  Query Match             21.0%;  Score 68;  DB 1;  Length 172;
  Best Local Similarity   35.0%;  Pred. No. 0.44;
  Matches   21;  Conservative    8;  Mismatches   23;  Indels    8;  Gaps    5;

Qy          2 MESGAP-ICHTCGEQVGH---DANGELFVACHECSY--PMCKSCFEFEIKEGRKVCLRCG 55
              |:|  | ||: |||  ||   | : :   ||: |     : | | | : :|  : |  ||
Db         40 MSSSLPDICYRCGES-GHLAKDCDLQEDKACYNCGRGGHIAKDCKEPK-REREQCCYNCG 97


RESULT 7
Z211_HUMAN
ID   Z211_HUMAN     STANDARD;      PRT;   367 AA.
AC   Q13398;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Zinc finger protein 211 (Zinc finger protein C2H2-25).
GN   ZNF211.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=97252692; PubMed=9095115;
RA   Becker K.G., Canning R.D., Nagle J.W., Dehejia A.M.,
RA   Polymeropoulos M.H., Lee I.J., Gado A.M., Biddison W.E., Drew P.D.;
```

```
T   "Molecular cloning and mapping of a novel developmentally regulated
T   human C2H2-type zinc finger.";
L   Mamm. Genome 8:287-289(1997).
N   [2]
P   PARTIAL SEQUENCE FROM N.A.
C   TISSUE=Hippocampus;
X   MEDLINE=95359976; PubMed=7633419;
A   Becker K.G., Nagle J.W., Canning R.D., Biddison W.E., Ozato K.,
A   Drew P.D.;
T   "Rapid isolation and characterization of 118 novel C2H2-type zinc
T   finger cDNAs expressed in human brain.";
L   Hum. Mol. Genet. 4:685-691(1995).
C   -!- FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR.
C   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
C   -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
C       FINGER PROTEINS.
C   -----------------------------------------------------------------------
C   This SWISS-PROT entry is copyright. It is produced through a collaboration
C   between the Swiss Institute of Bioinformatics and the EMBL outstation -
C   the European Bioinformatics Institute. There are no restrictions on its
C   use by non-profit institutions as long as its content is in no way
C   modified and this statement is not removed. Usage by and for commercial
C   entities requires a license agreement (See http://www.isb-sib.ch/announce/
C   or send an email to license@isb-sib.ch).
C   -----------------------------------------------------------------------
R   EMBL; U38904; AAA93261.1; -.
R   HSSP; P08045; 1ZNF.
R   Genew; HGNC:13003; ZNF211.
R   MIM; 601856; -.
R   GO; GO:0005634; C:nucleus; NAS.
R   GO; GO:0003700; F:transcription factor activity; NAS.
R   GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
R   InterPro; IPR007087; Znf_C2H2.
R   InterPro; IPR007086; Znf_C2H2_sub.
R   Pfam; PF00096; zf-C2H2; 11.
R   PRINTS; PR00048; ZINCFINGER.
R   ProDom; PD000003; Znf_C2H2; 7.
R   SMART; SM00355; ZnF_C2H2; 11.
R   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 11.
R   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 11.
W   Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
W   Nuclear protein; Repeat.
T   ZN_FING      60     82       C2H2-TYPE.
T   ZN_FING      88     110      C2H2-TYPE.
T   ZN_FING      116    138      C2H2-TYPE.
T   ZN_FING      144    166      C2H2-TYPE.
T   ZN_FING      172    194      C2H2-TYPE.
T   ZN_FING      200    222      C2H2-TYPE.
T   ZN_FING      228    250      C2H2-TYPE.
T   ZN_FING      256    278      C2H2-TYPE.
T   ZN_FING      284    306      C2H2-TYPE.
T   ZN_FING      312    334      C2H2-TYPE.
T   ZN_FING      340    362      C2H2-TYPE.
SQ  SEQUENCE   367 AA;  42154 MW;  AD6E99048F6719AE CRC64;

 Query Match           20.7%;  Score 67;  DB 1;  Length 367;
 Best Local Similarity  30.6%;  Pred. No. 1.2;
 Matches   19;  Conservative    5;  Mismatches   24;  Indels   14;  Gaps    4;

Qy          8 ICHTCGEQVGHDAN---------GELFVACHECSYPM-CKSCF--EFEIKEGRK--VCLR 53
              :|  ||:  | :|         ||  | | |||    |||     :  | :   !
Db        257 VCGECGKSFSHSSNLKNHQRVHTGERPVECSECSKSFSCKSNLIKHLRVHTGERPYECSE 316

Qy         54 CG 55
              ||
Db        317 CG 318


RESULT 8
ZN35_HUMAN
ID   ZN35_HUMAN     STANDARD;      PRT;   519 AA.
AC   P13682; Q96D01;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-OCT 1996 (Rel. 34, Last sequence update)
DT   15 SEP 2003 (Rel. 42, Last annotation update)
DE   Zinc finger protein 35 (Zinc finger protein HF.10).
GN   ZNF35
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92241871; PubMed=1572646;
RA   Lantrancone L., Pengue G., Pandolfi P.P., Salcini A.E., Giacomucci A.,
RA   Longo L., Donti E., de Sario A., la Mantia G., Pelicci P.G., Lania L.;
RT   "Structural and functional organization of the HF.10 human zinc
RT   finger gene (ZNF35) localized on chromosome 8q21-p22.";
RL   Genomics 12:720-728(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Placenta.
RX   MEDLINE=92241870; PubMed=1572645;
RA   Pannuti A., Pengue G., Pandolfi P.P., Salcini A.E., Giacomucci A.,
RA   Lania L.;
RT   "Isolation of cDNAs encoding finger proteins and measurement of the
RT   corresponding mRNA levels during myeloid terminal differentiation.";
RL   Nucleic Acids Res. 16:4277-4287(1988).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- FUNCTION: MAY BE INVOLVED IN TRANSCRIPTIONAL REGULATION. INVOLVED
CC       IN CELL DIFFERENTIATION AND/OR PROLIFERATION.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC   -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2 TYPE ZINC-
CC       FINGER PROTEINS.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; L35269; AAA85451.1; -.
DR   EMBL; M76702; AAA85451.1; JOINED.
DR   EMBL; M76703; AAA85451.1; JOINED.
DR   EMBL; X07289; CAA30268.1; -.
DR   EMBL; BC013597; AAH13597.1; -.
DR   PIR; A38073; A38073.
DR   HSSP; P08046; 1A1I.
DR   TRANSFAC; T04985; -.
DR   Genew; HGNC:13099; ZNF35.
DR   MIM; 194533; -.
```

```
DR   GO; GO:0003700; F:transcription factor activity; TAS.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; TAS.
DR   InterPro; IPR007087; Znf_C2H2.
DR   InterPro; IPR007086; Znf_C2H2_sub.
DR   Pfam; PF00096; zf-C2H2; 11.
DR   PRINTS; PR00048; ZINCFINGER.
DR   ProDom; PD000003; Znf_C2H2; 8.
DR   SMART; SM00355; ZnF_C2H2; 11.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 11.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 11.
KW   Transcription regulation; Zinc finger; DNA binding; Metal-binding;
KW   Nuclear protein; Repeat.
FT   DOMAIN          1    213       GLOBULAR DOMAIN.
FT   DOMAIN         13     19       POLY-GLU.
FT   ZN_FING       214    236       C2H2-TYPE.
FT   ZN_FING       242    264       C2H2-TYPE.
FT   ZN_FING       270    292       C2H2-TYPE.
FT   ZN_FING       298    320       C2H2 TYPE.
FT   ZN_FING       326    348       C2H2-TYPE.
FT   ZN_FING       354    376       C2H2-TYPE.
FT   ZN_FING       382    404       C2H2 TYPE.
FT   ZN_FING       410    432       C2H2-TYPE.
FT   ZN_FING       438    460       C2H2 TYPE.
FT   ZN_FING       466    488       C2H2-TYPE.
FT   ZN_FING       494    516       C2H2-TYPE.
FT   CONFLICT      405    432       MISSING (IN REF. 2)
FT   CONFLICT      424    424       S -> C (IN REF. 3)
SQ   SEQUENCE   519 AA;  58171 MW;  FB87F961CB3E61F C5764;

  Query Match             20.7%;  Score 67;  DB 1;  Length 519;
  Best Local Similarity   31.1%;  Pred. No. 1.6;
  Matches   19;  Conservative    4;  Mismatches   24;  Indels   14;  Gaps    3;

Qy          9 CHTCGEQVGHDAN---------GELFVACHECSYPMCKSC--FEFEIKEGRK--VCLRC 54
              |  ||:     ||         |:    |||:     :|           |:|  || :|
Db        216 CSVCGKGFSQSANLVVHQRIHTGEKPFECHECGKAFIQSANLVVHQRIHTGQKPYVCSKC 275

Qy         55 G 55
              |
Db        276 G 276


RESULT 9
RP3A_MOUSE
ID   RP3A_MOUSE     STANDARD;      PRT;   681 AA.
AC   P47709;
DT   01-FEB-1996 (Rel. 33, Created)
DT   28-FEB-2003 (Rel. 41, Last seq. update)
DT   28-FEB-2003 (Rel. 41, Last ann. update)
DE   Rabphilin-3A
GN   RPH3A.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6;
RA   Inagaki N.;
RL   Submitted (NOV-2001) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 1-606 FROM N.A.
RC   STRAIN=C57BL/6;
RX   MEDLINE=95122445; PubMed=7822236;
RA   Inagaki N., Mizuta M., Seino S.;
RT   "Cloning of a mouse Rabphilin-3A expressed in both neuroendocrine
RT   cells.";
RL   J. Biochem. 116:239-242(1994).
CC   -!- FUNCTION: PROTEIN TRANSPORT. PROBABLY INVOLVED WITH RAS-RELATED
CC       PROTEIN RAB-3A IN SYNAPTIC VESICLE TRAFFIC AND/OR SYNAPTIC
CC       VESICLE FUSION. COULD PLAY A ROLE IN NEUROTRANSMITTER RELEASE BY
CC       REGULATING MEMBRANE FLOW IN THE NERVE TERMINAL.
CC   -!- SUBUNIT: Monomer.
CC   -!- TISSUE SPECIFICITY: Specifically expressed in brain.
CC   -!- SIMILARITY: Contains 2 C2 domains.
CC   -!- SIMILARITY: Contains 1 FYVE-type zinc finger.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D29965; BAA06231.2; -.
DR   PIR; JX0338; JX0338.
DR   HSSP; P47709; 1ZBD.
DR   MGD; MGI:102788; Rph3a.
DR   InterPro; IPR000008; C2.
DR   InterPro; IPR002149; LRI.
DR   InterPro; IPR003315; RPH3A_effector.
DR   InterPro; IPR001565; Synaptotagmin.
DR   InterPro; IPR000306; Znf_FYVE.
DR   Pfam; PF00168; C2; 2.
DR   Pfam; PF02318; RPH3A_effector; 1.
DR   PRINTS; PR00360; C2DOMAIN.
DR   PRINTS; PR00399; SYNAPTOTAGMN.
DR   SMART; SM00239; C2; 2.
DR   PROSITE; PS00499; C2_DOMAIN_1; 2.
DR   PROSITE; PS50004; C2_DOMAIN_2; 2.
DR   PROSITE; PS50178; ZF_FYVE; 1.
KW   Repeat; Synapse; Protein transport; Zinc-finger.
FT   DOMAIN          1    274       RAB P25/SMG P25A BINDING (GTP GAMMA-S-
FT                                  BOUND).
FT   ZN_FING        88    145       FYVE-TYPE.
FT   DOMAIN        277    361       PRO-RICH.
FT   DOMAIN        381    485       C2 DOMAIN 1.
FT   DOMAIN        539    642       C2 DOMAIN 2.
SQ   SEQUENCE   681 AA;  75489 MW;  D09F8D8D2CBB271E CRC64;

  Query Match             20.7%;  Score 67;  DB 1;  Length 681;
  Best Local Similarity   32.7%;  Pred. No. 2.1;
  Matches   18;  Conservative    4;  Mismatches   23;  Indels   10;  Gaps    3;

Qy          5 GAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEIKEGR-------KVCL 52
              |   |  ||||:|    |   | | :|   :|  |   |     |       |:||
Db         90 GVNRCILCGEQLG--MLGSACVVCEDCKKNVCTKC-GVETSNNRPHPVWLCKICL 141


RESULT 10
RP3A_RAT
ID   RP3A_RAT       STANDARD;      PRT;   684 AA.
AC   P47709;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Rabphilin-3A.
GN   RPH3A.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95033210; PubMed=7946335;
RA   Li C., Takei K., Geppert M., Daniell L., Stenius K., Chapman E.R.,
RA   Jahn R., de Camilli P., Suedhof T.C.;
RT   "Synaptic targeting of rabphilin-3A, a synaptic vesicle
RT   Ca2+/phospholipid-binding protein, depends on rab3A/3C.";
RL   Neuron 13:885-898(1994).
RN   [2]
RP   X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF 44-167 IN COMPLEX WITH RAB3A.
RC   TISSUE=Brain;
RX   MEDLINE=99148269; PubMed=10025402;
RA   Ostermeier C., Brunger A.T.;
```

```
T   "Structural basis of Rab effector specificity: crystal structure of
T   the small G protein Rab3A complexed with the effector domain of
T   rabphilin-3A.";
L   Cell 96:363-374(1999).
C   -!- FUNCTION: PROTEIN TRANSPORT. PROBABLY INVOLVED WITH RAS-RELATED
C       PROTEIN RAB-3A IN SYNAPTIC VESICLE TRAFFIC AND/OR SYNAPTIC
C       VESICLE FUSION. COULD PLAY A ROLE IN NEUROTRANSMITTER RELEASE BY
C       REGULATING MEMBRANE FLOW IN THE NERVE TERMINAL.
C   -!- SUBUNIT: Monomer.
C   -!- TISSUE SPECIFICITY: Specifically expressed in brain.
C   -!- SIMILARITY: Contains 2 C2 domains.
C   -!- SIMILARITY: Contains 1 FYVE-type zinc finger.
C   ---------------------------------------------------------------------
C   This SWISS-PROT entry is copyright. It is produced through a collaboration
C   between the Swiss Institute of Bioinformatics and the EMBL outstation -
C   the European Bioinformatics Institute. There are no restrictions on its
C   use by non-profit institutions as long as its content is in no way
C   modified and this statement is not removed. Usage by and for commercial
C   entities requires a license agreement (See http://www.isb-sib.ch/announce/
C   or send an email to license@isb-sib.ch).
C   ---------------------------------------------------------------------
R   EMBL; U12571; AAA62662.1; -.
R   PIR; I58166; I58166.
R   PDB; 1ZBD; 12-APR-99.
R   PDB; 3RPB; 23-DEC-99.
R   InterPro; IPR000008; C2.
R   InterPro; IPR002149; LRI.
DR  InterPro; IPR003315; RPH3A_effector.
DR  InterPro; IPR001565; Synaptotagmin.
DR  InterPro; IPR000306; Znf_FYVE.
DR  Pfam; PF00168; C2; 2.
DR  Pfam; PF02318; RPH3A_effector; 1.
DR  PRINTS; PR00360; C2DOMAIN.
DR  PRINTS; PR00399; SYNAPTOTAGMN.
DR  SMART; SM00239; C2; 2.
DR  PROSITE; PS00499; C2_DOMAIN_1; 2.
DR  PROSITE; PS50004; C2_DOMAIN_2; 2.
DR  PROSITE; PS50178; ZF_FYVE; 1.
KW  Repeat; Synapse; Protein transport; Zinc-finger; 3D-structure.
FT  DOMAIN        1    277       RAB P25/SMG P25A BINDING (GTP GAMMA-S-
FT                               BOUND).
FT  ZN_FING      88    145       FYVE-TYPE.
FT  DOMAIN      280    364       PRO-RICH.
FT  DOMAIN      384    488       C2 DOMAIN 1.
FT  DOMAIN      542    645       C2 DOMAIN 2.
FT  HELIX        50     84
FT  TURN         85     85
FT  STRAND       93     93
FT  STRAND      100    100
FT  TURN        103    104
FT  STRAND      108    110
FT  TURN        112    114
FT  STRAND      117    119
FT  TURN        120    121
FT  STRAND      123    125
FT  STRAND      135    137
FT  HELIX       138    149
FT  TURN        150    150
FT  HELIX       152    155
FT  TURN        156    156
SQ  SEQUENCE   684 AA;  75832 MW;  05838BC3C7A86444 CRC64;

  Query Match             20.7%;  Score 67;  DB 1;  Length 684;
  Best Local Similarity   32.7%;  Pred. No. 2.1;
  Matches   18;  Conservative    4;  Mismatches   23;  Indels   10;  Gaps    3;

Qy          5 GAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEIKEGR-------KVCL 52
              |   | ||||:|    |   | | :|   :|  |   |     |       |:||
Db         90 GVNRCILCGEQLG--MLGSACVVCEDCKKNVCTKC-GVETSNNRPHPVWLCKICL 141


RESULT 11
HRX_MOUSE
```

```
ID  HRX_MOUSE      STANDARD;      PRT;  3666 AA.
AC  P55200;
DT  01-OCT-1996 (Rel. 34, Created)
DT  01-OCT-1996 (Rel. 34, Last sequence update)
DT  28-FEB-2003 (Rel. 41, Last annotation update)
DE  Zinc finger protein HRX (ALL-1) (Fragment).
GN  MLL OR HRX OR ALL1.
OS  Mus musculus (Mouse).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX  NCBI_TaxID=10090;
RN  [1]
RP  SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC  STRAIN=C57BL/6J, and C57BL/6 X CBA; TISSUE=Lung, and Spleen;
RX  MEDLINE=93317679; PubMed=8327517;
RA  Ma Q., Alder H., Nelson K.K., Chatterjee D., Gu Y., Nakamura T.,
RA  Canaani E., Croce C.M., Siracusa L.D., Buchberg A.M.;
RT  "Analysis of the murine All-1 gene reveals conserved domains with
RT  human ALL-1 and identifies a novel member with DNA
RT  methyltransferase.";
RL  Proc. Natl. Acad. Sci. U.S.A. 90:6350-6354(1993).
CC  -!- FUNCTION: Possibly acts as a transcriptional regulatory factor.
CC      May regulate genes involved in skeletal formation during embryonic
CC      development.
CC  -!- SUBUNIT: Interacts with SBF1 (By similarity).
CC  -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC  -!- ALTERNATIVE PRODUCTS:
CC      Event=Alternative splicing; Named isoforms=2;
CC      Name=1;
CC        IsoId=P55200-1; Sequence=Displayed;
CC      Name=2;
CC        IsoId=P55200-2; Sequence=VSP_006667;
CC  -!- SIMILARITY: BELONGS TO THE TRANSCRIPTION FACTOR TRITHORAX FAMILY
CC  -!- SIMILARITY: Contains 1 bromodomain.
CC  -!- SIMILARITY: Contains 1 SET domain.
CC  -!- SIMILARITY: Contains 3 PHD-type zinc fingers.
CC  -!- SIMILARITY: Contains 1 CXXC-type zinc finger.
CC  -!- SIMILARITY: Contains 1 post-SET domain.
CC  ---------------------------------------------------------------------
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  ---------------------------------------------------------------------
DR  EMBL; L17069; AAA62593.1;
DR  MGD; MGI:96995; MLL.
DR  GO; GO:0005634; C:nucleus; IDA.
DR  InterPro; IPR001487; Bromodomain.
DR  InterPro; IPR003889; FYrich_C.
DR  InterPro; IPR003888; FYrich_N.
DR  InterPro; IPR003616; PostSET.
DR  InterPro; IPR001214; SET.
DR  InterPro; IPR002857; Znf_CXXC.
DR  InterPro; IPR001965; Znf_PHD.
DR  Pfam; PF02178; AT_hook; 5.
DR  Pfam; PF00628; PHD; 3.
DR  Pfam; PF00856; SET; 1.
DR  Pfam; PF02008; zf-CXXC; 1.
DR  SMART; SM00297; BROMO; 1.
DR  SMART; SM00542; FYRC; 1.
DR  SMART; SM00541; FYRN; 1.
DR  SMART; SM00249; PHD; 4.
DR  SMART; SM00508; PostSET; 1.
DR  SMART; SM00317; SET; 1.
DR  PROSITE; PS50014; BROMODOMAIN_2; 1.
DR  PROSITE; PS50868; POST_SET; 1.
DR  PROSITE; PS50280; SET; 1.
DR  PROSITE; PS01359; ZF_PHD_1; 3.
DR  PROSITE; PS50016; ZF_PHD_2; 3.
KW  DNA-binding; Nuclear protein; Zinc-finger; Metal-binding;
```

```
KW   Transcription regulation; Bromodomain; Alternative splicing;
KW   Polymorphism.
FT   NON_TER       1      1
FT   DNA_BIND     67     78       A.T HOOK (BY SIMILARITY).
FT   DNA_BIND    115    125       A.T HOOK (BY SIMILARITY).
FT   DNA_BIND    199    207       A.T HOOK (BY SIMILARITY).
FT   ZN_FING    1044   1091       CXXC-TYPE.
FT   ZN_FING    1330   1381       PHD TYPE 1.
FT   ZN_FING    1383   1432       PHD TYPE 2.
FT   ZN_FING    1465   1529       PHD TYPE 3.
FT   DOMAIN     1605   1650       PROMODOMAIN (DIVERGENT).
FT   DOMAIN     3725   3846       SET.
FT   DOMAIN     3850   3866       POST-SET.
FT   DOMAIN       35     41       POLY-GLY.
FT   DOMAIN      459    469       PRO-RICH.
FT   DOMAIN     1231   1238       POLY-PRO.
FT   DOMAIN     3533   3536       POLY-GLU
FT   DOMAIN     3693   3697       POLY-GLU.
FT   VARSPLIC   1503   1505       Missing (in isoform 2).
FT                                /FTId=VSP_006667.
FT   VARIANT    1497   1497       K -> T.
SQ   SEQUENCE   3866 AA;  420976 MW;  ADFC55E14E8C6F1D CRC64;

  Query Match             20.7%;  Score 67;  DB 1;  Length 3866;
  Best Local Similarity   25.0%;  Pred. No. 11;
  Matches   16;  Conservative   10;  Mismatches   14;  Indels   24;  Gaps    4;

Qy          9 CHTCGE---------QVGHDANGELFVACHECS---------YPMCKSCFEFEIKEGRKVC 51
              | :||          |  ||      :  ||:|:         |:|  |:: :  | :
Db       1432 CKSCGSTTPGKGWDAQWSHD-----FSLCHDCAKLFAKGNFCPLCDKCYDDDDYESK--M 1484

Qy         52 LRCG 55
              ::||
Db       1485 MQCG 1488


RESULT 12
HRX_HUMAN
ID   HRX_HUMAN      STANDARD;      PRT;  3969 AA.
AC   Q03164; Q13743; Q13744; Q14845; Q16364; Q9UMA3;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Zinc finger protein HRX (ALL-1) (Trithorax-like protein).
GN   MLL OR HRX OR ALL1 OR TRX1 OR HTRX.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93046667; PubMed=1423624;
RA   Tkachuk D.C., Kohler S., Cleary M.L.;
RT   "Involvement of a homolog of Drosophila trithorax by 11q23
RT   chromosomal translocations in acute leukemias.";
RL   Cell 71:691-700(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96290553; PubMed=8703835;
RA   Nilson I., Loechner K., Siegler G., Greil J., Beck J.D., Fey G.H.,
RA   Marschalek R.;
RT   "Exon/intron structure of the human ALL-1 (MLL) gene involved in
RT   translocations to chromosomal region 11q23 and acute leukaemias.";
RL   Br. J. Haematol. 93:966-972(1996).
RN   [3]
RP   SEQUENCE OF 1-1909 FROM N.A.
RX   MEDLINE=93390935; PubMed=8378076;
RA   Yamamoto K., Seto M., Komatsu H., Iida S., Akao Y., Kojima S.,
RA   Kodera Y., Nakazawa S., Ariyoshi Y., Takahashi T., Ueda R.;
RT   "Two distinct portions of LTG19/ENL at 19p13 are involved in t(11;19)
RT   leukemia.";
RL   Oncogene 8:2617-2625(1993).
RN   [4]
RP   SEQUENCE OF 1317-2328 FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=93265134; PubMed=1303259;
RA   Djabali M., Selleri L., Parry P., Bower M., Young B.D., Evans G.A.;
RT   "A trithorax-like gene is interrupted by chromosome 11q23
RT   translocations in acute leukaemias.";
RL   Nat. Genet. 2:113-118(1992).
RN   [5]
RP   SEQUENCE OF 1251-1538 FROM N.A.
RX   MEDLINE=94215165; PubMed=8162575;
RA   Gu Y., Alder H., Nakamura T., Schichman S.A., Prasad R., Canaani O.,
RA   Saito H., Croce C.M., Canaani E.;
RT   "Sequence analysis of the breakpoint cluster region in the ALL-1 gene
RT   involved in acute leukemia.";
RL   Cancer Res. 54:2326-2330(1994).
RN   [6]
RP   SEQUENCE OF 1251-1654 FROM N.A. (ISOFORM 14P-18B).
RX   MEDLINE=95322025; PubMed=7598802;
RA   Mbangkollo D., Burnett R., McCabe N., Thirman M., Gill H., Yu H.,
RA   Rowley J.D., Diaz M.O.;
RT   "The human MLL gene: nucleotide sequence, homology to the Drosophila
RT   trx zinc-finger domain, and alternative splicing.";
RL   DNA Cell Biol. 14:475-483(1995).
RN   [7]
RP   SEQUENCE OF 1212-1603 FROM N.A.
RX   MEDLINE=95315013; PubMed=7794749;
RA   Marschalek R., Greil J., Lochner K., Nilson I., Siegler G.,
RA   Zweckbronner I., Beck J.D., Fey G.H.;
RT   "Molecular analysis of the chromosomal breakpoint and fusion
RT   transcripts in the acute lymphoblastic SEM cell line with chromosomal
RT   translocation t(4;11).";
RL   Br. J. Haematol. 90:308-320(1995).
RN   [8]
RP   SEQUENCE OF 1421-1540 FROM N.A.
RX   MEDLINE=94020842; PubMed=8414518;
RA   Forster A., Rabbitts T.H.;
RT   "A method for identifying genes within yeast artificial chromosomes:
RT   application to isolation of MLL fusion cDNAs from acute leukaemia
RT   translocations.";
RL   Oncogene 8:3157-3160(1993).
RN   [9]
RP   CHROMOSOMAL TRANSLOCATION WITH GAS7.
RX   MEDLINE=20183971; PubMed=10706619;
RA   Megonigal M.D., Cheung N.-K.V., Rappaport E.F., Nowell P.C.,
RA   Wilson R.B., Jones D.H., Addya K., Leonard D.G.B., Kushner B.H.,
RA   Williams T.M., Lange B.J., Felix C.A.;
RT   "Detection of leukemia-associated MLL-GAS7 translocation early during
RT   chemotherapy with DNA topoisomerase II inhibitors.";
RL   Proc. Natl. Acad. Sci. U.S.A. 97:2814-2819(2000).
RN   [10]
RP   CHROMOSOMAL TRANSLOCATION WITH AF3P21.
RX   MEDLINE=20115194; PubMed=10648423;
RA   Sano K., Hayakawa A., Piao J.-H., Kosaka Y., Nakamura H.;
RT   "Novel SH3 protein encoded by the AF3p21 gene is fused to the mixed
RT   lineage leukemia protein in a therapy-related leukemia with
RT   t(3;11)(p21;q23).";
RL   Blood 95:1066-1068(2000).
RN   [11]
RP   INTERACTION WITH SBF1.
RX   MEDLINE=98196760; PubMed=9537414;
RA   Cui X., De Vivo I., Slany R., Miyamoto A., Firestein R., Cleary M.L.;
RT   "Association of SET domain and myotubularin-related proteins modulates
RT   growth control.";
RL   Nat. Genet. 18:331-337(1998).
CC   -!- FUNCTION: Possibly acts as a transcriptional regulatory factor.
CC   -!- SUBUNIT: Interacts with SBF1.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=1;
CC         IsoId=Q03164-1; Sequence=Displayed;
CC       Name=14P-18B;
CC         IsoId=Q03164-2; Sequence=VSP_006666;
```

```
CC   -!- TISSUE SPECIFICITY: Heart, lung, brain and T and B lymphocytes.
CC   -!- DISEASE: Involved in acute leukemias by chromosomal translocations
CC       t(11;19)(q23;p13.3) that involves MLL and MLLT1/ENL;
CC       t(4;11)(q21;q23) that involves MLL and MLLT2/AF4; t(9;11)(p22;q23)
CC       that involves MLL and MLLT3/AF9; t(6;11)(q27;q23) that involves
CC       MLL and MLLT4/AF6; t(11;17)(q23;q21) that involves MLL and
CC       MLLT6/AF17; t(x;11)(q13;q23) that involves MLL and MLLT7/AFX1;
CC       t(10;11)(p12;q23) that involves MLL and MLLT10/AF10;
CC       t(1;11)(q21;q23) that involves MLL and AF1Q; t(11;19)(q23;p13.3)
CC       that involves MLL and ELL; t(11;19)(q23;p23) that involves MLL and
CC       GAS7; and t(3;11)(p21;q23) that involves MLL and AF3p21.
CC   -!- SIMILARITY: BELONGS TO THE TRANSCRIPTION FACTOR TRITHORAX FAMILY.
CC   -!- SIMILARITY: Contains 1 bromodomain.
CC   -!- SIMILARITY: Contains 1 SET domain.
CC   -!- SIMILARITY: Contains 3 PHD-type zinc fingers.
CC   -!- SIMILARITY: Contains 1 CXXC-type zinc finger.
CC   -!- SIMILARITY: Contains 1 post-SET domain.
CC   -!- DATABASE: NAME=Atlas Genet. Cytogenet. Oncol. Haematol.;
CC       WWW="http://www.infobiogen.fr/services/chromcancer/Genes/MLL.html".
CC   ----------------------------------- ------------------ ---------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; L04284; AAA58669.1; -.
DR   EMBL; Z69744; CAA93625.1; -.
DR   EMBL; Z69745; CAA93625.1; JOINED.
DR   EMBL; Z69746; CAA93625.1; JOINED.
DR   EMBL; Z69747; CAA93625.1; JOINED.
DR   EMBL; Z69748; CAA93625.1; JOINED.
DR   EMBL; Z69749; CAA93625.1; JOINED.
DR   EMBL; Z69750; CAA93625.1; JOINED.
DR   EMBL; Z69751; CAA93625.1; JOINED.
DR   EMBL; Z69752; CAA93625.1; JOINED.
DR   EMBL; Z69753; CAA93625.1; JOINED.
DR   EMBL; Z69754; CAA93625.1; JOINED.
DR   EMBL; Z69755; CAA93625.1; JOINED.
DR   EMBL; Z69756; CAA93625.1; JOINED.
DR   EMBL; Z69757; CAA93625.1; JOINED.
DR   EMBL; Z69758; CAA93625.1; JOINED.
DR   EMBL; Z69759; CAA93625.1; JOINED.
DR   EMBL; Z69760; CAA93625.1; JOINED.
DR   EMBL; Z69761; CAA93625.1; JOINED.
DR   EMBL; Z69762; CAA93625.1; JOINED.
DR   EMBL; Z69763; CAA93625.1; JOINED.
DR   EMBL; Z69764; CAA93625.1; JOINED.
DR   EMBL; Z69765; CAA93625.1; JOINED.
DR   EMBL; Z69766; CAA93625.1; JOINED.
DR   EMBL; Z69767; CAA93625.1; JOINED.
DR   EMBL; Z69768; CAA93625.1; JOINED.
DR   EMBL; Z69769; CAA93625.1; JOINED.
DR   EMBL; Z69770; CAA93625.1; JOINED.
DR   EMBL; Z69772; CAA93625.1; JOINED.
DR   EMBL; Z69773; CAA93625.1; JOINED.
DR   EMBL; Z69774; CAA93625.1; JOINED.
DR   EMBL; Z69775; CAA93625.1; JOINED.
DR   EMBL; Z69776; CAA93625.1; JOINED.
DR   EMBL; Z69777; CAA93625.1; JOINED.
DR   EMBL; Z69778; CAA93625.1; JOINED.
DR   EMBL; Z69779; CAA93625.1; JOINED.
DR   EMBL; Z69780; CAA93625.1; JOINED.
DR   EMBL; D14540; BAA03407.1; -.
DR   EMBL; L01986; AAA92511.1; -.
DR   EMBL; U04737; AAA18644.1; -.
DR   EMBL; S78570; AAB34770.1; -.
DR   EMBL; X83604; CAA58584.1; -.
DR   EMBL; S66432; AAB28545.1; -.
DR   EMBL; AF231998; AAG26332.2; ALT_TERM.
DR   PIR; A44265; A44265.
DR   TRANSFAC; T02337; -.
DR   Genew; HGNC:7132; MLL.
DR   MIM; 159555; -.
DR   GO; GO:0005634; C:nucleus; TAS.
DR   GO; GO:0003702; F:RNA polymerase II transcription factor acti...; TAS.
DR   GO; GO:0007048; P:oncogenesis; TAS.
DR   GO; GO:0006366; P:transcription from Pol II promoter; TAS.
DR   InterPro; IPR001487; Bromodomain.
DR   InterPro; IPR003889; FYrich_C.
DR   InterPro; IPR003888; FYrich_N.
DR   InterPro; IPR003616; Post_SET.
DR   InterPro; IPR001214; SET.
DR   InterPro; IPR002857; Znf_CXXC.
DR   InterPro; IPR001965; Znf_PHD.
DR   Pfam; PF02178; AT_hook; 2.
DR   Pfam; PF00628; PHD; 3.
DR   Pfam; PF00856; SET; 1.
DR   Pfam; PF02008; zf-CXXC; 1.
DR   SMART; SM00297; BROMO; 1.

  Query Match          .  %;  Score ..  , ...  Length 940;
  Best Local Similarity   2%;  Pred. No. 1.
  Matches   16;  Conservative   10;  Mismatches   14;  Indels   24;  Gaps    4;

Qy          9 CHTCGE-------- QVGHDANGELFVACHECS- ------YPMCKSCFEFEIKEGRKVC 51
              | :||          |  :|     :  | :  :           :|  |:: :  | :
Db       1533 CKSCGSTTPGKGWDAQWSHD     FSLCHE AKLFAKGNFCPLCDKCYDDDDYESK  M 1585

Qy         52 LRCG 55
              ::||
Db       1586 MQCG 1589


RESULT 13
MAT1_HUMAN
ID   MAT1_HUMAN     STANDARD;      PRT;   309 AA.
AC   P51948; Q15817;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   CDK-activating kinase assembly factor MAT1 (RING finger protein MAT1)
DE   (Menage a trois) (CDK7/cyclin H assembly factor) (p36) (p35) (Cyclin
DE   G1 interacting protein).
GN   MNAT1 OR MAT1 OR CAP35.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 23-48 AND 123-147.
RC   TISSUE=Placenta;
RX   MEDLINE=96091133; PubMed=8521818;
RA   Tassan J.-P., Jaquenoud M., Fry A.M., Frutiger S., Hughes G.J.,
RA   Nigg E.A.;
RT   "In vitro assembly of a functional human CDK7-cyclin H complex
RT   requires MAT1, a novel 36 kDa RING finger protein.";
RL   EMBO J. 14:5608-5617(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Cervical carcinoma;
RX   MEDLINE=96105021; PubMed=8521393;
RA   Yee A., Nichols M., Wu E., Hall F.L., Kobiyashi R., Xiong Y.;
RT   "Molecular cloning of CDK7-associated human MAT1, a cyclin-dependent
RT   kinase-activating kinase (CAK) assembly factor.";
RL   Cancer Res. 55:6058-6062(1995).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Xu F., Hall F.L., Starnes V., Wu L.;
RL   Submitted (JUN-1996) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Rieder M.J., Livingston R.J., Daniels M.R., Montoya M.A., Chung M.-W.,
RA   Miyamoto K.E., Nguyen C.P., Nguyen D.A., Poel C.L., Robertson P.D.,
```

```
RA   Schackwitz W.S., Sherwood J.K., Witrak L.A., Nickerson D.A.;
RL   Submitted (OCT-2002) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Placenta;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [6]
RP   FUNCTION.
RX   MEDLINE=99149021; PubMed=10024882;
RA   Tirode F., Busso D., Coin F., Egly J.-M.;
RT   "Reconstitution of the transcription factor TFIIH: assignment of
RT   functions for the three enzymatic subunits, XPB, XPD, and cdk7.";
RL   Mol. Cell 3:87-95(1999).
CC   -!- FUNCTION: Stabilizes the cyclin H-CDK7 complex to form a
CC       functional CDK-activating kinase (CAK) enzymatic complex. CAK
CC       activates the cyclin-associated kinases CDC2/CDK1, CDK2, CDK4 and
CC       CDK6 by threonine phosphorylation. CAK complexed to the core-TFIIH
CC       basal transcription factor activates RNA polymerase II by serine
CC       phosphorylation of the repetitive carboxyl terminus domain (CTD)
CC       of its large subunit (POLR2A), allowing its escape from the
CC       promoter and elongation of the transcripts. Involved in cell cycle
CC       control and in RNA transcription by RNA polymerase II.
CC   -!- SUBUNIT: Associates primarily with CDK7 and cyclin H to form the
CC       CAK complex. CAK can further associate with the core TFIIH to
CC       form the TFIIH basal transcription factor.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- TISSUE SPECIFICITY: HIGHEST LEVELS IN [illegible] MODERATE
CC       LEVELS ARE PRESENT THYMUS, [illegible] VARY, AS SMALL INTESTINE
CC       THE LOWEST LEVELS ARE FOUND IN [illegible]
CC   -!- SIMILARITY: Contains 1 RING-type zinc finger.
CC   -!- SIMILARITY: Contains 1 ubiquitin interacting motif (UIM) repeat.
CC   ---------- ---------- ---------- ----------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------- ---------- ---------- ----------
DR   EMBL; X97843; CAA61112.1; -.
DR   EMBL; X92669; CAA63356.1; -.
DR   EMBL; U61835; AAB05248.1; -.
DR   EMBL; AY165512; AAN47195.1; -.
DR   EMBL; BC000820; AAH00820.1; -.
DR   PIR; S60157; S60157.
DR   PDB; 1G25; 04-FEB-03.
DR   TRANSFAC; T02187; -.
DR   Genew; HGNC:7181; MNAT1.
DR   GK; P51948; -.
DR   MIM; 602659; -.
DR   GO; GO:0008270; F:zinc ion binding; TAS.
DR   GO; GO:0006281; P:DNA repair; TAS.
DR   GO; GO:0006461; P:protein complex assembly; TAS.
DR   GO; GO:0000079; P:regulation of CDK activity; TAS.
DR   GO; GO:0006357; P:regulation of transcription from Pol II pro. . .; TAS.
DR   InterPro; IPR004575; MAT1.
DR   InterPro; IPR003903; UIM.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00184; RING; 1.
DR   TIGRFAMs; TIGR00570; cdk7; 1.
DR   PROSITE; PS50330; UIM; 1.
DR   PROSITE; PS00518; ZF_RING_1; 1.
DR   PROSITE; PS50089; ZF_RING_2; 1.
KW   Transcription regulation; Cell cycle; Nuclear protein; Zinc-finger;
KW   3D-structure.
FT   ZN_FING       6     50       RING-TYPE.
FT   DOMAIN      142    161       UIM.
FT   CONFLICT    188    229       MISSING (IN REF. 3).
FT   CONFLICT    247    248       EY -> DN (IN REF. 3).
SQ   SEQUENCE   309 AA;  35823 MW;  6818DDE230E81A97 CRC64;

  Query Match             20.5%;  Score 66.5;  DB 1;  Length 309;
  Best Local Similarity   26.8%;  Pred. No. 1.1;
  Matches   15;  Conservative    9;  Mismatches   27;  Indels    5;  Gaps    1;

Qy          1 MMESGAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEIKEGRKVCLRCGS 56
              |  :  |  |  |  |    :      |  |  :  :  |  :|:||  :      |   |  ||:
Db          1 MDDQGCPRCKTTKYR-----NPSLKLMVNVCGHTLCESCVDLLFVRGAGNCPECGT 51


RESULT 14
MAT1_MOUSE
ID   MAT1_MOUSE     STANDARD;      PRT;   309 AA.
AC   P51949; Q9D8A0; Q9D8D2;
DT   01-OCT-1996 (Rel. 34, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   CDK-activating kinase assembly factor MAT1 (RING finger protein MAT1)
DE   (Menage a trois) (CDK7/cyclin H assembly factor) (p36) (p35).
GN   MNAT1 OR MAT1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96006524; PubMed=7553872;
RA   Fisher R.P., Jin P., Chamberlin H.M., Morgan D.O.;
RT   "Alternative mechanisms of CAK assembly require an assembly factor or
RT   an activating kinase.";
RL   Cell 83:47-57(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Small intestine;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
```

```
L   Nature 409:685-690(2001).
C   -!- FUNCTION: Stabilizes the cyclin H-CDK7 complex to form a
C       functional CDK-activating kinase (CAK) enzymatic complex. CAK
C       activates the cyclin-associated kinases CDC2/CDK1, CDK2, CDK4 and
C       CDK6 by threonine phosphorylation. CAK complexed to the core-TFIIH
C       basal transcription factor activates RNA polymerase II by serine
C       phosphorylation of the repetitive carboxyl-terminus domain (CTD)
C       of its large subunit (POLR2A), allowing its escape from the
C       promoter and elongation of the transcripts. Involved in cell cycle
C       control and in RNA transcription by RNA polymerase II.
C   -!- SUBUNIT: Associates primarily with CDK7 and cyclin H to form the
C       CAK complex. CAK can further associate with the core-TFIIH to
C       form the TFIIH basal transcription factor.
C   -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
C   -!- SIMILARITY: Contains 1 RING-type zinc finger.
C   -!- SIMILARITY: Contains 1 ubiquitin-interacting motif (UIM) repeat.
C   -----------------------------------------------------------------------
C   This SWISS-PROT entry is copyright. It is produced through a collaboration
C   between the Swiss Institute of Bioinformatics and the EMBL outstation -
C   the European Bioinformatics Institute. There are no restrictions on its
C   use by non-profit institutions as long as its content is in no way
C   modified and this statement is not removed. Usage by and for commercial
C   entities requires a license agreement (See http://www.isb-sib.ch/announce/
C   or send an email to license@isb-sib.ch).
CC  -----------------------------------------------------------------------
DR  EMBL; U35249; AAA91741.1; -.
DR  EMBL; AK008249; BAB25556.1; -.
DR  EMBL; AK008134; BAB25483.1; -.
DR  PIR; A57235; A57235.
DR  MGD; MGI:106207; Mnat1.
DR  InterPro; IPR004575; MAT1.
DR  InterPro; IPR003903; UIM.
DR  InterPro; IPR001841; Znf_ring.
DR  Pfam; PF00097; zf-C3HC4; 1.
DR  TIGRFAMs; TIGR00570; cdk7; 1.
DR  PROSITE; PS50330; UIM; 1.
DR  PROSITE; PS00518; ZF_RING_1; 1.
DR  PROSITE; PS50089; ZF_RING_2; 1.
KW  Transcription regulation; Cell cycle; Nuclear protein; Zinc-finger.
FT  ZN_FING       6     50       RING-TYPE.
FT  DOMAIN      142    161       UIM.
FT  CONFLICT     15     15       R -> P (IN REF. 2; BAB25483).
FT  CONFLICT     74     74       I -> F (IN REF. 2; BAB25483).
FT  CONFLICT    117    117       N -> K (IN REF. 2; BAB25483).
FT  CONFLICT    139    140       KL -> NV (IN REF. 2; BAB25483).
FT  CONFLICT    162    162       R -> Q (IN REF. 1).
SQ  SEQUENCE   309 AA;  35848 MW;  C72CD43175D69499 CRC64;

  Query Match             20.5%;  Score 66.5;  DB 1;  Length 309;
  Best Local Similarity   26.8%;  Pred. No. 1.1;
  Matches   15;  Conservative    9;  Mismatches   27;  Indels    5;  Gaps    1;

Qy          1 MMESGAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEIKEGRKVCLRCGS 56
              |  :  | | | |    :      |  | :  :  | : :|:|| :     |   |  ||:
Db          1 MDDQGCPRCKTTKYR-----NPSLKLMVNVCGHTLCESCVDLLFVRGAGNCPECGT 51


RESULT 15
LIK1_HUMAN
ID   LIK1_HUMAN     STANDARD;      PRT;   647 AA.
AC   P53667; O15283; Q15820; Q15821; Q9Y5Q1;
DT   01-OCT-1996 (Rel. 34, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   LIM domain kinase 1 (EC 2.7.1.37) (LIMK-1).
GN   LIMK1 OR LIMK.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RX   MEDLINE=94239736; PubMed=8183554;
RA   Mizuno K., Okano T., Tatsuki R., Nakano K., Kuma K.-I., Miyata T.,
RA   Nakamura T.;
RT   "Identification of a brain cDNA encoding a novel protein kinase with
RT   two repeats of the LIM/double zinc finger motif.";
RL   Oncogene 9:1605-1612(1994).
RN   [2]
RP   SEQUENCE FROM N.A. [ISOFORMS 1 AND 2].
RC   TISSUE=Hippocampus, and Placenta;
RX   MEDLINE=96291399; PubMed=8689688;
RA   Frangiskakis J.M., Ewart A.K., Morris C.A., Mervis C.B.,
RA   Bertrand J., Robinson B.F., Klein B.P., Ensing G.J., Everett L.A.,
RA   Green E.D., Proschel C., Gutowski N.J., Noble M., Atkinson D.L.,
RA   Odelberg S.J., Keating M.T.;
RT   "LIM-kinase1 hemizygosity implicated in impaired visuospatial
RT   constructive cognition.";
RL   Cell 86:59-69(1996).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96411691; PubMed=8811549;
RA   Osborne L.R., Martindale D.W., Scherer S.W., Shi X.-M., Hutzenga J.,
RA   Heng H.H.Q., Tsui L.-C.;
RA   Koop B.F., Tsui L.-C.;
RT   "Identification of genes from a 5-cM contig at 7q11.23 that is
RT   commonly deleted in Williams syndrome patients.";
RL   Genomics 36:328-336(1996).
RN   [4]
RP   SEQUENCE FROM N.A. (ISOFORM 3).
RC   TISSUE=Epidermal carcinoma;
RX   MEDLINE=99214220; PubMed=10196247;
RA   Edwards D.C., Gill G.N.;
RT   "Structural features of LIM kinase that control effects on the actin
RT   cytoskeleton.";
RL   J. Biol. Chem. 274:11352-11361(1999).
CC   -!- FUNCTION: May be a component of an intracellular signaling pathway
CC       and may be involved in brain development. Phosphorylates and
CC       inactivates the actin binding/depolymerizing factor cofilin and
CC       induces actin cytoskeletal changes. Isoform 3 has a dominant
CC       negative effect on actin cytoskeletal changes.
CC   -!- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.
CC   -!- SUBUNIT: THE LIM DOMAIN INTERACTS WITH THE CYTOPLASMIC DOMAIN OF
CC       PRO-NEUREGULIN-1 (PRO-NRG1).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=3;
CC       Name=1;
CC         IsoId=P53667-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=P53667-2; Sequence=VSP_003125;
CC       Name=3;
CC         IsoId=P53667-3; Sequence=VSP_003126, VSP_003127;
CC   -!- TISSUE SPECIFICITY: HIGHEST EXPRESSION IN BOTH ADULT AND FETAL
CC       NERVOUS SYSTEM. DETECTED UBIQUITOUSLY THROUGHOUT THE DIFFERENT
CC       REGIONS OF ADULT BRAIN, WITH HIGHEST LEVELS IN THE CEREBRAL
CC       CORTEX. EXPRESSED TO A LESSER EXTENT IN HEART AND SKELETAL MUSCLE.
CC   -!- DISEASE: Haploinsufficiency of LIMK1 may be the cause of certain
CC       cardiovascular and musculo-skeletal abnormalities observed in
CC       Williams-Beuren syndrome (WBS), a rare developmental disorder. It
CC       is a contiguous gene deletion syndrome involving genes from
CC       chromosome band 7q11.23.
CC   -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC   -!- SIMILARITY: Contains 2 LIM zinc binding domains.
CC   -!- SIMILARITY: Contains 1 PDZ/DHR domain.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; D26309; BAA05371.1; -.
DR   EMBL; U62293; AAB17545.1; -.
```

```
DR   EMBL; U62293; AAB17546.1; -.
DR   EMBL; U63721; AAC13885.1; -.
DR   EMBL; U63721; AAC13886.1; -.
DR   EMBL; AF134379; AAD25742.1; -.
DR   PIR; JP0078; JP0078.
DR   HSSP; Q05158; 1QLI.
DR   Genew; HGNC:6613; LIMK1.
DR   MIM; 601329; -.
DR   MIM; 194050; -.
DR   GO; GO:0005737; C:cytoplasm; TAS
DR   GO; GO:0004674; F:protein serine/threonine kinase activity; TAS.
DR   GO; GO:0007012; P:actin cytoskeleton reorganization; TAS.
DR   GO; GO:0007399; P:neurogenesis; TAS.
DR   GO; GO:0006468; P:protein amino acid phosphorylation; TAS.
DR   GO; GO:0007266; P:Rho protein signal transduction; TAS.
DR   InterPro; IPR001781; LIM.
DR   InterPro; IPR001478; PDZ.
DR   InterPro; IPR000719; Prot_kinase
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   Pfam; PF00412; LIM; 2.
DR   Pfam; PF00595; PDZ; 1.
DR   Pfam; PF00069; pkinase; 1.
DR   ProDom; PD000094; LIM; 2.
DR   ProDom; PD000001; Prot_kinase; 1.
DR   SMART; SM00132; LIM; 2.
DR   SMART; SM00228; PDZ; 1.
DR   PROSITE; PS00478; LIM_DOMAIN_1; 2.
DR   PROSITE; PS50023; LIM_DOMAIN_2; 2.
DR   PROSITE; PS50106; PDZ; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; FALSE_NEG.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transferase; Serine/threonine-protein kinase; ATP binding; Repeat;
KW   LIM domain; Metal-binding; Zinc; Williams-Beuren syndrome;
KW   Alternative splicing.
FT   DOMAIN       25     75       LIM 1.
FT   DOMAIN       84    137       LIM 2.
FT   DOMAIN      165    258       PDZ.
FT   DOMAIN      339    604       PROTEIN KINASE.
FT   NP_BIND     345    353       ATP (BY SIMILARITY).
FT   BINDING     368    368       ATP (BY SIMILARITY).
FT   ACT_SITE    460    460       BY SIMILARITY.
FT   VARSPLIC      1     14       Missing (in isoform 2).
FT                                /FTId=VSP_003125.
FT   VARSPLIC    294    305       LRSSIDRSFGA -> PARTWVALSFSA (in isoform
FT                                3).
FT                                /FTId=VSP_003126.
FT   VARSPLIC    506    647       Missing (in isoform 3).
FT                                /FTId=VSP_003127.
FT   CONFLICT    580    580       Y -> F (IN REF. 1).
SQ   SEQUENCE   647 AA;  72601 MW;  5B48BEB6659076C5 CRC64;

  Query Match             20.4%;  Score 56;  DB 1;  Length 647;
  Best Local Similarity   18.9%;  Pred. No. 2.6;
  Matches   21;  Conservative   14;  Mismatches   12;  Indels   64;  Gaps    5;

Qy          1 MMESGA--PICHTCGEQV--------   ---   ----------  - -GHDANGELFV-- 26
              |  |  |:  |:|  :|||:::                          :: :|:||
Db         15 MGEEGSELPVCASCGQRIYDGQYLQALNADWHADCFRCCDCSASLSHQYYEKDGQLFCKK 74

Qy         27 - ---   ACHECSYPMCK----  -    - SCFEFEIKEGRKVCLRCGS 56
                     :|| ||  : |            !!          :,| ||:
Db         75 DYWARYGESCHGCSEQITKGLVMVAGELKYHPECF------   ICLTCGT 116


Search completed: October 29, 2003, 18:29:35
Job time : 3.19149 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 29, 2003, 18:26:33 ; Search time 1.78723 Seconds
(without alignments)
3013.285 Million cell updates/sec

Title: US-09-980-043-4
Perfect score: 324
Sequence: 1 MMESGAPICHTCGEQVGHDA.....EKSCFEPFIKEGRKVCLRCGS 56

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_76:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No	Score	% Query Match	Length	DB	ID	Description
1	264	81.5	974	2	T10797	cellulose synthase
2	246	75.9	958	2	T04870	cellulose synthase
3	161	49.7	326	2	T08581	TGACG-motif bindin
4	161	49.7	583	2	T02[illegible]	cellulose synthase
5	155	47.8	1065	2	F84[illegible]	probable cellulose
6	154	47.5	322	2	T08[illegible]	TGACG motif-bindin
7	152	46.9	1026	2	T51[illegible]	cellulose synthase
8	152	46.9	1081	2	T05[illegible]	cellulose synthase
9	149	46.0	322	2	T12[illegible]	TGACG-motif bindin
10	148	45.7	1084	2	T08583	cellulose synthase
11	144	44.4	1088	2	H84604	probable cellulose
12	139	42.9	1081	2	T52028	cellulose synthase
13	138	42.6	1065	2	T52054	cellulose synthase
14	78	24.1	292	2	T32000	hypothetical prote
15	76.5	23.6	1145	2	T51546	cellulose synthase
16	74	22.8	560	2	S53382	protein YKR029c ho
17	71.5	22.1	190	2	A82370	DNA topoisomerase
18	71.5	22.1	271	2	T40238	hypothetical prote
19	71.5	22.1	1958	2	T39808	hypothetical prote
20	69	21.3	697	2	H86497	78.1K hypothetical
21	68.5	21.1	73	2	T28185	hypothetical prote
22	68.5	21.1	267	2	G02784	cyclin G1 interact
23	68.5	21.1	950	2	F86222	hypothetical prote
24	68.5	21.1	1353	2	T19157	probable metal bin
25	68.5	21.1	1424	2	T19156	probable metal bin
26	67	20.7	366	2	I53035	trithorax homolog
27	67	20.7	519	2	A38073	transcription acti
28	67	20.7	681	2	JX0318	rabphilin-3A - mou
29	67	20.7	684	2	I58166	rabphilin-3A - rat
30	67	20.7	3869	2	A48205	All-1 protein +GTE
31	67	20.7	3968	2	A44265	trithorax homolog
32	66.5	20.5	309	1	S60157	RING finger protei
33	66.5	20.5	309	1	A57235	RING finger protei
34	66	20.4	647	1	JP0078	LIM protein kinase
35	65.5	20.2	314	2	JC2324	LIM protein - huma
36	65.5	20.2	329	2	T20546	hypothetical prote
37	65.5	20.2	546	2	T32382	hypothetical prote
38	65.5	20.2	739	2	A83015	primosomal protein
39	65.5	20.2	1927	2	T25604	hypothetical prote
40	65	20.1	491	2	S52920	disintegrin (EC 3.
41	65	20.1	647	1	I48737	LIM protein kinase
42	65	20.1	704	2	A48097	rabphilin-3A - bov
43	65	20.1	733	2	D97112	primosomal protein
44	64.5	19.9	186	2	S11569	hypothetical prote
45	64.5	19.9	309	2	S60120	RING finger protei

ALIGNMENTS

RESULT 1
T10797
cellulose synthase (EC 2.4.1.-) catalytic chain celA1 - upland cotton
C;Species: Gossypium hirsutum (upland cotton)
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
C;Accession: T10797
R;Pear, J.R.; Kawagoe, Y.; Schreckengost, W.E.; Delmer, D.P.; Stalker, D.M.
Proc. Natl. Acad. Sci. U.S.A. 93, 12637-12642, 1996
A;Title: Higher plants contain homologs of the bacterial celA genes encoding the cata
A;Reference number: Z17152; MUID:97057296; PMID:8901635
A;Accession: T10797
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-974 <PEA>
A;Cross-references: EMBL:U58283; NID:g1706955; PIDN:AAB37766.1; PID:g1706956
A;Experimental source: strain Acala SJ-2; fiber
C;Genetics:
A;Gene: celA1
C;Function:
A;Description: involved in the synthesis of cellulose
C;Keywords: glycosyltransferase; hexosyltransferase

Query Match 81.5%; Score 264; DB 2; Length 974;
Best Local Similarity 75.0%; Pred. No. 7.3e-19;
Matches 42; Conservative 8; Mismatches 6; Indels 0; Gaps 0;

Qy 1 MMESGAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEIKEGRKVCLRCGS 56
Db 1 MMESGVPVCHTCGEHVGLNVNGEPFVACHECNFPICKSCFEYDLKEGRKACLRCGS 56

RESULT 2
T04870
cellulose synthase (EC 2.4.1.-) catalytic chain F28A21.190 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-Jun-1999
C;Accession: T04870
R;Bevan, M.; Mueller, M.W.; Muendlein, A.; Felber, R.; Bancroft, I.; Mewes, H.W.; May
submitted to the Protein Sequence Database, February 1999
A;Reference number: Z15387
A;Accession: T04870
A;Molecule type: DNA
A;Residues: 1-958 <BEV>
A;Cross-references: EMBL:AL035526
A;Experimental source: cultivar Columbia; BAC clone F28A21
C;Genetics:
A;Map position: 4
A;Introns: 60/1; 76/3; 94/1; 149/2; 240/2; 355/3; 401/3; 443/3; 514/3; 662/3; 780/3;
A;Note: F28A21.190
C;Keywords: glycosyltransferase; hexosyltransferase

Query Match 75.9%; Score 246; DB 2; Length 958;

```
 Best Local Similarity   67.9%;  Pred. No. 4.4e-17;
 Matches   38;  Conservative   12;  Mismatches    6;  Indels    0;  Gaps    0;

y           1 MMESGAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEIKEGRKVCLRCGS 56
              |||| :|||:||||::|  :||| |||||||||:|:||:| |:| ||||::|||||:
b           1 MMESRSPICNTCGEEIGVKSNGEFFVACHECSFPICKACLEYEFKEGRRICLRCGN 56


ESULT 3
08591
GACG-motif binding protein STF1 - soybean
;Species: Glycine max (soybean)
;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 11-Jun-1999
;Accession: T08591
;Cheong, Y.H.; Yoo, C.M.; Park, J.M.; Ryu, G.R.; Goekjian, V.H.; Nagao, R.T.; Key, J.L.
ubmitted to the EMBL Data Library, September 1995
;Description: STF1 is a novel TGACG-binding factor with a zinc-finger motif and a bZIP
;Reference number: Z16445
;Accession: T08591
;Status: preliminary; translated from GB/EMBL/DDBJ
;Molecule type: mRNA
;Residues: 1-326 <CHE>
;Cross-references: EMBL:L28003; NID:g2934883; PID:g2934884
;Experimental source: strain Williams; hypocotyl

 Query Match             49.7%;  Score 161;  DB 2;  Length 326;
 Best Local Similarity   51.1%;  Pred. No. 5.8e-09;
 Matches   24;  Conservative   10;  Mismatches   13;  Indels    0;  Gaps    0;

y           8 ICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEIKEGRKVCLRC 54
              ||  ||: :|  | |:||||||| :|:| ||:|:|:|   : | :|
b          38 ICQICGDTIGLTATGDLFVACHECGFPLCHSCYEYELKNVSQSCPQC 84


ESULT 4
02209
ellulose synthase (EC 2.4.1.-) catalytic chain - rice (fragment)
;Species: Oryza sativa (rice)
;Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 11-Jun-1999
;Accession: T02209
;Arioli, T.; Peng, L.; Betzner, A.S.; Burn, J.; Wittke, W.; Herth, W.; Camilleri, C.; H
cience 279, 717-720, 1998
;Title: Molecular analysis of cellulose biosynthesis in Arabidopsis.
;Reference number: Z13745; MUID:98111412; PMID:9445479
;Accession: T02209
;Status: translated from GB/EMBL/DDBJ
;Molecule type: mRNA
;Residues: 1-583 <ARI>
;Cross-references: EMBL:AF030052; NID:g2781432; PID:g2781433
;Experimental source: subsp. Japonica, strain Nipponbare
;Keywords: glycosyltransferase; hexosyltransferase

 Query Match             49.7%;  Score 161;  DB 2;  Length 583;
 Best Local Similarity   48.9%;  Pred. No. 8.8e-09;
 Matches   23;  Conservative   13;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          8 ICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEIKEGRKVCLRC 54
              :|  ||: ||  | |::||||:||::|:|: |:|:| ||| : | :|
Db         40 VCQICGDTVGVSATGDVFVACNECAFPVCRPCYEYERKEGNQCCPQC 86


RESULT 5
F84649
robable cellulose synthase catalytic subunit [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: F84649
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: F84649
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1065 <STO>
A;Cross-references: GB:AE002093; NID:g4447865; PIDN:AAD20713.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g25540
A;Map position: 2

  Query Match             47.8%;  Score 155;  DB 2;  Length 1065;
  Best Local Similarity   48.4%;  Pred. No. 5.3e-08;
  Matches   23;  Conservative   11;  Mismatches   13;  Indels    0;  Gaps    0;

Qy          8 ICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEIKEGRKVCLRC 54
              |:  ||: ||    | :: ||:   :|.:||: :' |:| : | :|
Db         32 ICQICGDEVGLTKDNVFVACNECGFPLCQSCYEYERKDGSQCCPQC 78


RESULT 6
T08592
TGACG-motif-binding protein STF2 - soybean
C;Species: Glycine max (soybean)
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 11-Jun-1999
C;Accession: T08592
R;Cheong, Y.H.; Yoo, C.M.; Park, J.M.; Ryu, G.R.; Goekjian, V.H.; Nagao, R.T.; Key, J.L.
submitted to the EMBL Data Library, September 1995
A;Description: STF1 is a novel TGACG-binding factor with a zinc-finger motif and a bZIP
A;Reference number: Z16445
A;Accession: T08592
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-322 <CHE>
A;Cross-references: EMBL:L28004; NID:g934886; PID:g2934885
A;Experimental source: strain Williams; hypocotyl
C;Genetics:
A;Gene: STF2

  Query Match             47.8%;  Score 154;  DB 2;  Length 322;
  Best Local Similarity   47.8%;  Pred. No. 2.9e-08;
  Matches   22;  Conservative   11;  Mismatches   13;  Indels    0;  Gaps    0;

Qy          9 CHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEIKEGRKVCLRC 54
              |  ||: :;  | |::|||.||| :|:|   :|:|:    : | :|
Db         39 CQICGDTIGLTATGDVFVACHECGFPLCHSCYEYELKHMSQSCPQC 84


RESULT 7
T51579
cellulose synthase catalytic subunit (IRX3) - Arabidopsis thaliana
N;Alternate names: protein T10B6_80
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 18-Aug-2000
C;Accession: T51579
R;Sato, S.; Nakamura, Y.; Kaneko, T.; Kato, T.; Asamizu, E.; Kotani, H.; Tabata, S.; Mew
submitted to the Protein Sequence Database, August 2000
A;Reference number: Z25394
A;Accession: T51579
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1026 <SAT>
A;Cross-references: EMBL:AL391142
A;Experimental source: cultivar Columbia; BAC clone T10B6
C;Genetics:
A;Map position: 5
A;Introns: 25/3; 91/1; 184/1; 215/2; 304/2; 429/3; 507/3; 578/3; 650/1; 715/3; 833/3
A;Note: T10B6_80

  Query Match             46.9%;  Score 152;  DB 2;  Length 1026;
  Best Local Similarity   47.8%;  Pred. No. 1e-07;
  Matches   22;  Conservative   11;  Mismatches   13;  Indels    0;  Gaps    0;

Qy          9 CHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEIKEGRKVCLRC 54
              |  ||:|:|    ::|||||:|| :| .: |:|:: :|| : | :|
```

```
Db          37 CEICGDQIGLTVEGDLFVACNECGFPACRPCYEYERREGTQNCPQC 82


RESULT 8
T05351
cellulose synthase (EC 2.4.1.-) catalytic chain RSW1 - Arabidopsis thaliana
N;Alternate names: protein F8B4.110
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-Jun-1999
C;Accession: T05351
R;Bevan, M.; Terryn, N.; Ardiles, W.; Huysshaert, C.; Lasseville, R.; De Clerck, R.; De
ewes, H.W.; Mayer, K.F.X.; Schueller, C.
submitted to the Protein Sequence Database, February 1999
A;Reference number: Z15409
A;Accession: T05351
A;Molecule type: DNA
A;Residues: 1-1081 <BEV>
A;Cross-references: EMBL:AL034567
A;Experimental source: cultivar Columbia; BAC clone F8B4
C;Genetics:
A;Gene: RSW1
A;Map position: 4
A;Introns: 27/3; 93/1; 150/3; 189/1; 252/2; 341/2; 456/3; 532/3; 544/3; 615/3; 704/1; 76
A;Note: F8B4.110
C;Keywords: glycosyltransferase; hexosyltransferase; P loop

  Query Match             46.9%;  Score 152;  DB 2;  Length 1081;
  Best Local Similarity   46.8%;  Pred. No. 1.1e-07;
  Matches   22;  Conservative   13;  Mismatches   12;  Indels    0;  Gaps    0;

Qy           8 ICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEIKEGRKVCLRC 54
               ||  ||: ||    |::|||:||::  :|. |:|:| |:|   | ::
Db          38 ICQICGDDVGLAETGDVFVACNECAFPVCRPCYEYERKDGTQ CPQC 84


RESULT 9
T12093
TGACG-motif binding protein - fava bean
C;Species: Vicia faba (fava bean)
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 20-Jun-2000
C;Accession: T12093
R;Wohlfarth, T.
submitted to the EMBL Data Library, May 1996
A;Reference number: Z17415
A;Accession: T12093
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-522 <WOH>
A;Cross-references: EMBL:X97904

  Query Match             46.0%;  Score 14[illegible];  DB 2;  Length [illegible];
  Best Local Similarity   47.8%;  Pred. No. 0.[illegible];
  Matches   22;  Conservative   11;  Mismatches   [illegible];  Indels    0;  Gaps    0;

Qy           9 CHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEIKEGRKVCLRC 54
               !  ||: :|  | | :||||:||::::  |:::|||  : : :|
Db          36 CQICGDTIGLTAAGGVFVACHECAFPLCHPCYEYEIKNVSLCPQC 81


RESULT 10
T08583
cellulose synthase (EC 2.4.1.-) catalytic chain - Arabidopsis thaliana
N;Alternate names: protein T22F8.250
C;Species: Arabidopsis thaliana (mouse ear cress)
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 22-Oct-1999
C;Accession: T08583; T09014
R;Bevan, M.; Zimmermann, W.; Gruenheisen, A.; Wambutt, R.; Bancroft, I.; Mewes, H.W.; May
submitted to the Protein Sequence Database, May 1999
A;Reference number: Z16442
A;Accession: T08583
A;Molecule type: DNA
A;Residues: 1-1094 <BEV>
A;Cross-references: EMBL:AL050351; GSPDB:GN00014; ATSP:T22F8.250
A;Experimental source: cultivar Columbia; BAC clone T22F8
R;Arioli, T.; Peng, L.; Betzner, A.S.; Burn, J.; Wittke, W.; Herth, W.; Camilleri, C.
Science 279, 717-720, 1998
A;Title: Molecular analysis of cellulose biosynthesis in Arabidopsis.
A;Reference number: Z13745; MUID:98111412; PMID:9445479
A;Accession: T09014
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1084 <ARI>
A;Cross-references: EMBL:AF027173; NID:g2827140; PIDN:AAC39335.1; PID:g2827141
A;Experimental source: cultivar Columbia
C;Genetics:
A;Gene: ATSP:T22F8.250; Ath-A
A;Map position: 4
A;Introns: 27/3; 156/3; 191/1; 254/2; 343/2; 458/3; 504/3; 546/3; 617/3; 708/1; 773/3.
C;Keywords: cell wall synthesis; glycosyltransferase; hexosyltransferase

  Query Match             45.7%;  Score 148;  DB 2;  Length 1084;
  Best Local Similarity   38.9%;  Pred. No. 2.7e-07;
  Matches   21;  Conservative   15;  Mismatches   18;  Indels    0;  Gaps    0;

Qy           1 MMESGAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEIKEGRKVCLRC 54
               : |     |  ||:::    : ||||||:||::|:|: |:|:|  :|| : | :|
Db          31 VQELSGQTCQICGDEIELTVSSELFVACNECAFPVCRPCYEYERREGNQACPQC 84


RESULT 11
H84604
probable cellulose synthase catalytic subunit [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: H84604
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: H84604
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1088 <STO>
A;Cross-references: GB:AE002093; NID:g4417271; PIDN:AAD20396.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g21770
A;Map position: 2

  Query Match             44.4%;  Score 144;  DB 2;  Length 1088;
  Best Local Similarity   38.9%;  Pred. No. 6.7e-07;
  Matches   21;  Conservative   14;  Mismatches   19;  Indels    0;  Gaps    0;

Qy           3 ESGAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEIKEGRKVCLRCGS 56
               |     |  | :::    ||| |:||:||::| |: |:|:| :|| : | :||:
Db          33 ELSGQTCKICRDEIELTDNGEPFIACNECAFPTCRPCYEYERREGNQACPQCGT 86


RESULT 12
T52028
cellulose synthase [imported] - Arabidopsis thaliana (fragment)
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 20-Oct-2000
C;Accession: T52028
R;Joshi, C.
submitted to the EMBL Data Library, May 1998
A;Reference number: Z25890
A;Accession: T52028
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1081 <JOS>
A;Cross-references: EMBL:AF062485; PIDN:AAC29067.1

  Query Match             42.9%;  Score 139;  DB 2;  Length 1081;
  Best Local Similarity   37.0%;  Pred. No. 2.1e-06;
```

```
 Matches   20;  Conservative   15;  Mismatches   19;  Indels    0;  Gaps    0;

Qy          1 MMESGAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEIKEGRKVCLRC 54
              : |     |  | ::    :|| ||||:||::|:|: |:|:| :|: : | :|
Db         28 VQELSGQTCQICRDEIELTVDGEPFVACNECAFPVCRPCYEYERREGNQACPQC 81


RESULT 13
T52054
cellulose synthase (EC 2.4.1.-) catalytic subunit [validated]  Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 31 Dec 2000
C;Accession: T52054
R;Arioli, T.; Peng, L.; Betzner, A.S.; Burn, J.; Wittke, W.; Herth, W.; Camilleri, C.; H
Science 279, 717-720, 1998
A;Title: Molecular analysis of cellulose biosynthesis in Arabidopsis.
A;Reference number: Z13745; MUID:98111412; PMID:9445479
A;Accession: T52054
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1065 <ARI>
A;Cross-references: EMBL:AF027174; PIDN:AAC39336.1
C;Genetics:
A;Note: Ath-B
C;Function:
A;Description: EC 2.4.1.-; cellulose synthase [validated, MUID:98111412]; involved in as
C;Keywords: glycosyltransferase; hexosyltransferase

  Query Match             42.6%;  Score 138;  DB 2;  Length 1065;
  Best Local Similarity   38.5%;  Pred. No. 2.6e-06;
  Matches   25;  Conservative   12;  Mismatches   16;  Indels   12;  Gaps    2;

Qy          2 MES-----GAPI-------CHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEIKEGRK 49
              |||     | |:       |  | : ||   :|: ||||  ||:|:|: |:|:| |:| :
Db          1 MESEGETAGKPMKNIVPQTCQICSDNVGKTVDGDRFVACDICSFPVCRPCYEYERKDGNQ 60

Qy         50 VCLRC 54
              | :|
Db         61 SCPQC 65


RESULT 14
T32000
hypothetical protein B0281.8 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 02-Sep-2000
C;Accession: T32000
R;Pauley, A.; Scheet, P.
submitted to the EMBL Data Library, July 1997
A;Description: The sequence of C. elegans cosmid B0281.
A;Reference number: Z21109
A;Accession: T32000
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-292 <PAU>
A;Cross-references: EMBL:AF016666; PIDN:AAB66086.1; GSPDB:GN00020; CESP:B0281.8
A;Experimental source: strain Bristol N2; clone B0281
C;Genetics:
A;Gene: CESP:B0281.8
A;Map position: 2
A;Introns: 27/1; 85/3; 108/3; 171/2; 186/3; 270/1
C;Superfamily: RING finger homology
F;2-57/Domain: RING finger homology <RRN>

  Query Match             24.1%;  Score 78;  DB 2;  Length 292;
  Best Local Similarity   29.2%;  Pred. No. 0.96;
  Matches   14;  Conservative   10;  Mismatches   24;  Indels    0;  Gaps    0;

Qy          7 PICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEIKEGRKVCLRC 54
              | |  | :    : :|:      :||: :|: | :  :| || ||  |
Db          4 PKCEICDDDFSSEEDGDHNPRNLKCSHTLCEGCIKKLLKNGRVVCPFC 51


RESULT 15
T51546
cellulose synthase catalytic subunit [validated] protein - Arabidopsis thaliana
N;Alternate names: protein F2K13.60
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 18-Aug-2000
C;Accession: T51546
R;Sato, S.; Nakamura, Y.; Kaneko, T.; Katoh, T.; Asamizu, E.; Kotani, H.; Tabata, S.; New
submitted to the Protein Sequence Database, August 2000
A;Reference number: Z25194
A;Accession: T51546
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1145 <SAT>
A;Cross-references: EMBL:AL[illegible]1141
A;Experimental source: cultivar Columbia; BAC clone F2K13
C;Genetics:
A;Map position: 5
A;Introns: 297/2; 566/2
A;Note: F2K13.60

  Query Match             [illegible]  Score 138;  DB 2;  Length 1145;
  Best Local Similarity   [illegible]
  Matches   14;  Conservative   8;  Mismatches   20;  Indels    1;  Gaps    1;

Qy         12 CGEQVGHDANGELFVACHECSYPMCKSCFEFEIKEGRKVCLRC 54
              |  :   |  |:  : | || : :  :|  :| |  :|  |
Db        133 CDAKVMSDERGQD LPC ECDFKIKPRKFIFAVKTGGGICPGC 176


Search completed: October [illegible], 2003, 15:31:33
Job time : 3.78723 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 31, 2003, 17:42:48 ; Search time 42.7853 Seconds
(without alignments)
3559.830 Million cell updates/sec

Title: US-09-980-043-4
Perfect score: 324
Sequence: 1 MMESGAPICHTCGEQVGHDA.....KSCFEFEIKEGRKVCLRCGS 56

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1811591 seqs, 1359896290 residues

Total number of hits satisfying chosen parameters: 3623182

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09980043/runat_29102003_163854_21959/app_query.fasta_1.1358
DB=Published_Applications_NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 END=-1 -MATRIX=blosum62
-TRANS human40.cdi -LIST=45 -DOCALIGN=200 THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
MAXLEN=2000000000 -USER=US09980043_@CGN_1_1_350_&runat_29102003_163854_21959
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=-1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	264	81.5	3328	10	US-09-838-539-1	Sequence 1, Appli
2	164	50.6	3563	10	US-09-900-237-25	Sequence 25, Appl
3	162	50.0	3753	12	US-10-160-719-33	Sequence 33, Appl
4	162	50.0	3753	12	US-10-160-719-53	Sequence 53, Appl
5	162	50.0	3780	12	US-10-209-059-1	Sequence 1, Appli
6	162	50.0	3780	12	US-10-160-719-9	Sequence 9, Appli
7	161	49.7	1741	14	US-10-229-193-13	Sequence 13, Appl
8	161	49.7	3725	12	US-10-209-059-45	Sequence 45, Appl
9	161	49.7	3725	12	US-10-160-719-13	Sequence 13, Appl
10	161	49.7	3725	12	US-10-160-719-21	Sequence 21, Appl
11	161	49.7	3725	12	US-10-160-719-41	Sequence 41, Appl
12	158	48.8	3568	12	US-10-209-059-41	Sequence 41, Appl
13	158	48.8	3568	12	US-10-160-719-1	Sequence 1, Appli
14	158	48.8	3776	10	US-09-900-237-3	Sequence 3, Appli
15	154	47.5	3517	10	US-09-900-237-13	Sequence 13, Appl
16	153.5	47.4	3786	10	US-09-900-237-7	Sequence 7, Appli
17	153.5	47.4	3813	12	US-10-209-059-17	Sequence 17, Appl
18	153.5	47.4	3813	12	US-10-160-719-25	Sequence 25, Appl
19	153.5	47.4	3813	12	US-10-160-719-45	Sequence 45, Appl
20	152	46.9	3603	14	US-10-229-193-5	Sequence 5, Appli
21	152	46.9	3673	14	US-10-229-193-11	Sequence 11, Appl
22	152	46.9	8411	14	US-10-229-193-3	Sequence 3, Appli
23	150	46.3	3936	10	US-09-900-237-9	Sequence 9, Appli
24	150	46.3	3969	12	US-10-209-059-49	Sequence 49, Appl
25	150	46.3	3969	12	US-10-160-719-17	Sequence 17, Appl
26	150	46.3	3969	12	US-10-160-719-37	Sequence 37, Appl
27	148	45.7	3255	10	US-09-938-842A-559	Sequence 559, App
28	148	45.7	3746	12	US-10-160-719-29	Sequence 29, Appl
29	148	45.7	3746	12	US-10-160-719-49	Sequence 49, Appl
30	148	45.7	3773	12	US-10-209-059-9	Sequence 9, Appli
31	148	45.7	3773	12	US-10-160-719-5	Sequence 5, Appli
32	148	45.7	3799	12	US-10-267-459-5	Sequence 5, Appli
33	148	45.7	3799	12	US-10-209-059-21	Sequence 21, Appl
34	148	45.7	3828	14	US-10-229-193-7	Sequence 7, Appli
35	148	45.7	5009	14	US-10-229-193-4	Sequence 4, Appli
36	146	45.1	3704	12	US-10-209-059-13	Sequence 13, Appl
37	146	45.1	3704	12	US-10-160-719-57	Sequence 57, Appl
38	140	43.2	641	9	US-09-770-149-611	Sequence 611, App
39	139	42.9	3255	10	US-09-938-842A-2530	Sequence 2530, Ap
40	138	42.6	3470	12	US-10-209-059-25	Sequence 25, Appl
41	138	42.6	3614	14	US-10-229-193-9	Sequence 9, Appli
42	135	41.7	3626	10	US-09-900-237-29	Sequence 29, Appl
43	115.5	35.6	3231	12	US-10-209-059-27	Sequence 27, Appl
44	89	27.5	203	10	US-09-924-035A-317	Sequence 317, App
45	87	26.9	1063	10	US-09-838-539-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-09-838-539-1
; Sequence 1, Application US/09838539
; Patent No. US20020129401A1
; GENERAL INFORMATION:
; APPLICANT: Stalker, D. et al.
; TITLE OF INVENTION: Plant Cellulose Synthase and Promoter
; TITLE OF INVENTION: Sequences
; FILE REFERENCE: 15621/03/US
; CURRENT APPLICATION NUMBER: US/09/838,539
; CURRENT FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/029,987
; PRIOR FILING DATE: 1996-10-29
; PRIOR APPLICATION NUMBER: 08/960,048
; PRIOR FILING DATE: 1997-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 3328
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Oligonucleotide
US-09-838-539-1

```
lignment Scores:
red. No.:               8.19e-28        Length:         3328
core:                   264.00          Matches:        42
ercent Similarity:      89.29%          Conservative:   8
est Local Similarity:   75.00%          Mismatches:     6
uery Match:             81.48%          Indels:         0
B:                      10              Gaps:           0

S-09-980-043-4 (1-56) x US-09-838-539-1 (1-3328)

y          1 MetMetGluSerGlyAlaProIleCysHisThrCysGlyGluGlnValGlyHisAspAla 20
             ||||||||||||||    |||:::|||||||||||||||||||    ||||||    :::
b        179 ATGATGGAATCTGGGGTTCCTGTTTGCCACACTTGTGGTGAACATGTTGGGTTGAATGTT 238

y         21 AsnGlyGluLeuPheValAlaCysHisGluCysSerTyrProMetCysLysSerCysPhe 40
             |||||||||   ||||||||||||||||||||||:::::|||:::|||||||||||||||
b        239 AATGGTGAACCTTTTGTGGCTTGCCATGAATGTAATTTCCCTATTTGTAAGAGTTGTTTT 298

y         41 GluPheGluIleLysGluGlyArgLysValCysLeuArgCysGlySer 56
             |||:::::::::|||||||||||||||    ||||:|||||||||||||
b        299 GAGTATGATCTTAAGGAAGGACGAAAAGCTTGCTTGCGTTGTGGTAGT 346

ESULT 2
S-09-900-237-25
 Sequence 25, Application US/09900237
 Patent No. US20020120124A1
 GENERAL INFORMATION:
  APPLICANT: Allen, Stephen
  TITLE OF INVENTION: Plant Cellulose Synthases
  FILE REFERENCE: BB1170 US CIP
  CURRENT APPLICATION NUMBER: US/09/900,237
  CURRENT FILING DATE:  2001-07-06
  PRIOR APPLICATION NUMBER: 60/092,844
  PRIOR FILING DATE: 1998-07-14
  PRIOR APPLICATION NUMBER: PCT/US99/15871
  PRIOR FILING DATE: 1999-07-13
  PRIOR APPLICATION NUMBER: 09/720383
  PRIOR FILING DATE: 2000-12-21
  NUMBER OF SEQ ID NOS: 33
  SOFTWARE: Microsoft Office 97
 SEQ ID NO 25
   LENGTH: 3563
   TYPE: DNA
   ORGANISM: Impatiens balsamia
S-09-900-237-25

lignment Scores:
red. No.:               2.45e-13        Length:         3563
core:                   164.00          Matches:        24
ercent Similarity:      68.52%          Conservative:   13
est Local Similarity:   44.44%          Mismatches:     17
uery Match:             50.62%          Indels:         0
B:                      10              Gaps:           0

US-09-980-043-4 (1-56) x US-09-900-237-25 (1-3563)

Qy         1 MetMetGluSerGlyAlaProIleCysHisThrCysGlyGluGlnValGlyHisAspAla 20
             :::   |||:::          ||||||      |||||||:::   ||||||     |||
Db        97 CTAAAAGAGGCTAACGGACAGATATGTCAGATATGCGGTGACACAGTCGGAAAATCAGCC 156

Qy        21 AsnGlyGluLeuPheValAlaCysHisGluCysSerTyrProMetCysLysSerCysPhe 40
             |||:::   ||||||||||||:::||||||   :::||:::|||:::   |||:::
Db       157 ACCGGCGACACTTTCGTTGCCTGTAATGAATGTGGATTCCCCGTTTGCCGGCCTTGTTAC 216

Qy        41 GluPheGluIleLysGluGlyArgLysValCysLeuArgCys 54
             |||:::|||   |||:::|||   :::   |||   :::|||
Db       217 GAGTACGAAAGGAAAGATGGAAACCAATGCTGCCCTCAGTGC 258

ESULT 3
US-10-160-719-33
 Sequence 33, Application US/10160719
 Publication No. US20030167528A1
 GENERAL INFORMATION:
 APPLICANT: Kanwarpal S. Dhugga
 APPLICANT:  Timothy G. Helentjaris
 APPLICANT:  Benjamin A.  Powen
 APPLICANT:  Xun Wang
 TITLE OF INVENTION: Maize Cellulose Synthases and Uses
 TITLE OF INVENTION:  Thereof
 FILE REFERENCE: 0864
 CURRENT APPLICATION NUMBER: US/10/160,719
 CURRENT FILING DATE:  2002-06-03
 PRIOR APPLICATION NUMBER: US/09/321,393A
 PRIOR FILING DATE: 1999-05-26
 PRIOR APPLICATION NUMBER: US 60/096,822
 PRIOR FILING DATE: 1998-08-17
 NUMBER OF SEQ ID NOS: 60
 SOFTWARE: FastSEQ for Windows Version 3.0
SEQ ID NO 33
  LENGTH: 3753
  TYPE: DNA
  ORGANISM: Zea mays
  FEATURE:
  NAME/KEY: CDS
  LOCATION: (184)...(3408)
US-10-160-719-33

Alignment Scores:
Pred. No.:              5.1e-13         Length:         3753
Score:                  162.00          Matches:        23
Percent Similarity:     76.67%          Conservative:   13
Best Local Similarity:  48.94%          Mismatches:     11
Query Match:            50.00%          Indels:         0
DB:                     12              Gaps:           0

US-09-980-043-4 (1-56) x US-10-160-719-33 (1-3753)

Qy         8 IleCysHisThrCysGlyGluGlnValGlyHisAspAlaAsnGlyGluLeuPheValAla 27
             :::|||      : |||:::   :|||        |||    ||:::::||:|||||
Db       301 GTCTGCCAGATTTGCGGTGACTCTGTGGGTGTTTCAGCCACTGGTGATGTCTTTGTTGCC 360

Qy        28 CysHisGluCysSer TyrProMetCysLysSerCysPheGluPheGluIleLysGluGly 47
             |||:::|:||||:: :::|||::: |||:::   |||:::|| :::|||  ||.|||||
Db       361 TGCAATGAGTGTGGC TTCCCTGTCTGCCGCCATGCTATGAGTATGAGCGCAAGGAGGGG 420

Qy        48 ArgLysValCysLeuArgCys 54
                :::   ||   ::: :|
Db       421 AACCAATGCTGCCC CAGTGC 441

RESULT 4
US-10-160-719-53
; Sequence 53, Application US/10160719
; Publication No. US20030167528A1
; GENERAL INFORMATION:
;  APPLICANT: Kanwarpal S. Dhugga
;  APPLICANT:  Timothy G. Helentjaris
;  APPLICANT:  Benjamin A.  Powen
;  APPLICANT:  Xun Wang
;  TITLE OF INVENTION: Maize Cellulose Synthases and Uses
;  TITLE OF INVENTION:  Thereof
;  FILE REFERENCE: 0864
;  CURRENT APPLICATION NUMBER: US/10/160,719
;  CURRENT FILING DATE:  2002-06-03
;  PRIOR APPLICATION NUMBER: US/09/321,393A
;  PRIOR FILING DATE: 1999-05-26
;  PRIOR APPLICATION NUMBER: US 60/096,822
;  PRIOR FILING DATE: 1998-08-17
;  NUMBER OF SEQ ID NOS: 60
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 53
;   LENGTH: 3753
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (184)...(3408)
```

US 10-160-719-53

Alignment Scores:
Pred. No.: 5.1e-13 Length: 3753
Score: 162.00 Matches: 23
Percent Similarity: 76.60% Conservative: 13
Best Local Similarity: 48.94% Mismatches: 11
Query Match: 50.00% Indels: 0
DB: 12 Gaps: 0

US-09-980-043-4 (1-56) x US-10-160-719-53 (1-3753)

```
Qy          8 IleCysHisThrCysGlyGluGlnValGlyHisAspAlaAsnGlyGluLeuPheValAla 27
              :::|||      ||||||:::   ||||||      |||   |||:::::|||||||||
Db        301 GTCTGCCAGATTTGCGGTGACTCTGTGGGTGTTTCAGCCACTGGTGATGTCTTTGTTGCC 360

Qy         28 CysHisGluCysSerTyrProMetCysLysSerCysPheGluPheGluIleLysGluGly 47
              |||:::||||||:::::|||:::|||:::   |||:::|| :::| |   :|||||||
Db        361 TGCAATGAGTGTGCCTTCCCTGTCTGCCGCCCATGCTATGAGTATGAGCGCAAGGAGGGG 420

Qy         48 ArgLysValCysLeuArgCys 54
              :::   |||   :::|||
Db        421 AACCAATGCTGCCCCCAGTGC 441
```

RESULT 5
US 10-209-059-1
; Sequence 1, Application US/10209059
; Publication No. US20030163838A1
; GENERAL INFORMATION:
; APPLICANT: Dhugga, Kanwarpal S.
; APPLICANT: Wang, Haiyin
; TITLE OF INVENTION: Maize Cellulose Synthases and Uses
; TITLE OF INVENTION: Thereof
; FILE REFERENCE: 0864R2
; CURRENT APPLICATION NUMBER: US/10/209,059
; CURRENT FILING DATE: 2002-07-31
; PRIOR APPLICATION NUMBER: 60/096,822
; PRIOR FILING DATE: 1998-08-17
; PRIOR APPLICATION NUMBER: 09/371,383
; PRIOR FILING DATE: 1999-08-06
; PRIOR APPLICATION NUMBER: 09/550,488
; PRIOR FILING DATE: 2000-04-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 3780
; TYPE: DNA
; ORGANISM: Zea mays
US-10-209-059-1

Alignment Scores:
Pred. No.: 5.15e-13 Length: 3780
Score: 162.00 Matches: 23
Percent Similarity: 76.60% Conservative: 13
Best Local Similarity: 48.94% Mismatches: 11
Query Match: 50.00% Indels: 0
DB: 12 Gaps: 0

US-09-980-043-4 (1-56) x US-10-209-059-1 (1-3780)

```
Qy          8 IleCysHisThrCysGlyGluGlnValGlyHisAspAlaAsnGlyGluLeuPheValAla 27
              :::|||      ||||||:::   |||:|      |||   |||:::::|||||||||
Db        318 GTCTGCCAGATTTGCGGTGACTCTGTGGGTGTTTCAGCCACTGGTGATGTCTTTGTTGCC 377

Qy         28 CysHisGluCysSerTyrProMetCysLysSerCysPheGluPheGluIleLysGluGly 47
              |||:::||||||:::::.||:::.|||:::   |: :::| :::|||   ||||:||||
Db        378 TGCAATGAGTGTGCCTTCCCTGTCTGCCGCCCATGCTATGAGTATGAGCGCAAGGAGGGG 437

Qy         48 ArgLysValCysLeuArgCys 54
              :::   |||   :::|||
Db        438 AACCAATGCTGCCCCCAGTGC 458
```

RESULT 6

US-10-160-719-9
; Sequence 9, Application US/10160719
; Publication No. US20030167528A1
; GENERAL INFORMATION:
; APPLICANT: Kanwarpal S. Dhugga
; APPLICANT: Timothy G. Helentjaris
; APPLICANT: Benjamin A. Bowen
; APPLICANT: Xun Wang
; TITLE OF INVENTION: Maize Cellulose Synthases and Uses
; TITLE OF INVENTION: Thereof
; FILE REFERENCE: 0864
; CURRENT APPLICATION NUMBER: US/10/160,719
; CURRENT FILING DATE: 2002-06-03
; PRIOR APPLICATION NUMBER: US/09/371,383A
; PRIOR FILING DATE: 1999-08-06
; PRIOR APPLICATION NUMBER: US 60/096,822
; PRIOR FILING DATE: 1998-08-17
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 9
; LENGTH: 3780
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (201)...(3425)
US-10-160-719-9

Alignment Scores:
Pred. No.: 5.15e-13 Length: 3780
Score: 162.00 Matches: 23
Percent Similarity: 76.60% Conservative: 13
Best Local Similarity: 48.94% Mismatches: 11
Query Match: 50.00% Indels: 0
DB: 12 Gaps: 0

US-09-980-043-4 (1-56) x US-10-160-719-9 (1-3780)

```
Qy          8 IleCysHisThrCysGlyGluGlnValGlyHisAspAlaAsnGlyGluLeuPheValAla 27
              :::|||      ||||||:::   ||||||      |||   |||:::::|||||||||
Db        318 GTCTGCCAGATTTGCGGTGACTCTGTGGGTGTTTCAGCCACTGGTGATGTCTTTGTTGCC 377

Qy         28 CysHisGluCysSerTyrProMetCysLysSerCysPheGluPheGluIleLysGluGly 47
              |||:::||||||:::::.| |:::|||:::   |||:::|||:::|||   ||||||||
Db        378 TGCAATGAGTGTGCCTTCCCTGTCTGCCGCCCATGCTATGAGTATGAGCGCAAGGAGGGG 437

Qy         48 ArgLysValCysLeuArgCys 54
              :::   |||   :::|||
Db        438 AACCAATGCTGCCCCCAGTGC 458
```

RESULT 7
US-10-229-193-13
; Sequence 13, Application US/10229193
; Publication No. US20030126643A1
; GENERAL INFORMATION:
; APPLICANT: Arioli, Antonio Williamson, Richard E.
; APPLICANT: Betzner, Andreas S.
; APPLICANT: Peng, Liangcai
; TITLE OF INVENTION: Manipulation of Cellulose and/or Beta-1.4-Glucan
; FILE REFERENCE: 96-98A
; CURRENT APPLICATION NUMBER: US/10/229,193
; CURRENT FILING DATE: 2002-08-26
; PRIOR APPLICATION NUMBER: 09/221,013
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: PCT/AU97/00402
; PRIOR FILING DATE: 1997-06-24
; PRIOR APPLICATION NUMBER: AU PO0699
; PRIOR FILING DATE: 1996-06-27
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 1741
; TYPE: DNA

```
   ORGANISM: Oryza sativa
   FEATURE:
   NAME/KEY: CDS
   LOCATION: (101)..(1741)
S-10-229-193-13

lignment Scores:
red. No.:                2.64e-13        Length:          1741
core:                    161.00          Matches:           23
ercent Similarity:       76.60%          Conservative:      13
est Local Similarity:    48.94%          Mismatches:        11
uery Match:              49.69%          Indels:             0
B:                       14              Gaps:               0

S-09-980-043-4 (1-56) x US-10-229-193-13 (1-1741)

y           8 IleCysHisThrCysGlyGluGlnValGlyHisAspAlaAsnGlyGluLeuPheValAla 27
              :::|||      ||||||:::   ||||||      |||   |||:::::::|||||||||
b         218 GTCTGCCAGATTTGTGGCGACACTGTTGGCGTCTCGGCCACCGGCGACGTCTTTGTTGCC 277

y          28 CysHisGluCysSerTyrProMetCysLysSerCysPheGluPheGluIleLysGluGly 47
              |||:::||||||:::::::|||:::|||:::   |||:::|||:::|||   ||||||||
b         278 TGCAATGAGTGCGCCTTCCCGGTCTGCCGCCCTTGCTACGAGTACGAACGCAAGGAAGGG 337

y          48 ArgLysValCysLeuArgCys 54
                 :::   |||   :::|||
b         338 AACCAGTGCTGCCCCCAGTGC 358

ESULT 8
S-10-209-059-45
 Sequence 45, Application US/10209059
 Publication No. US20030163838A1
 GENERAL INFORMATION:
  APPLICANT: Dhugga, Kanwarpal S.
  APPLICANT:  Wang, Haiyin
  TITLE OF INVENTION: Maize Cellulose Synthases and Uses
  TITLE OF INVENTION:  Thereof
  FILE REFERENCE: 0864R2
  CURRENT APPLICATION NUMBER: US/10/209,059
  CURRENT FILING DATE:  2002-07-31
  PRIOR APPLICATION NUMBER: 60/096,822
  PRIOR FILING DATE: 1998-08-17
  PRIOR APPLICATION NUMBER: 09/371,383
  PRIOR FILING DATE: 1999-08-06
  PRIOR APPLICATION NUMBER: 09/550,483
  PRIOR FILING DATE: 2000-04-14
  NUMBER OF SEQ ID NOS: 52
  SOFTWARE: FastSEQ for Windows Version 4.0
 SEQ ID NO 45
   LENGTH: 3725
   TYPE: DNA
   ORGANISM: Zea mays
S-10-209-059-45

lignment Scores:
red. No.:                7.04e-13        Length:          3725
core:                    161.00          Matches:           23
ercent Similarity:       76.60%          Conservative:      13
est Local Similarity:    48.94%          Mismatches:        11
uery Match:              49.69%          Indels:             0
B:                       12              Gaps:               0

S-09-980-043-4 (1-56) x US-10-209-059-45 (1-3725)

y           8 IleCysHisThrCysGlyGluGlnValGlyHisAspAlaAsnGlyGluLeuPheValAla 27
              :::|||      ||||||:::   ||||||      |||   |||::::::|||||||||
b         296 GTCTGCCAGATTTGTGGCGACACTGTTGGCGTTTCAGCCACTGGTGATGTCTTTGTTGCC 355

y          28 CysHisGluCysSerTyrProMetCysLysSerCysPheGluPheGluIleLysGluGly 47
              |||:::||||||:::::::|||:::|||:::   |||:::|||:::|||   ||||||||
b         356 TGCAATGAGTGTGCCTTCCCTGTCTGCCGCCCTTGCTATGAGTACGAGCGCAAGGAAGGG 415

y          48 ArgLysValCysLeuArgCys 54
```

```
                :::   |   ::|||
Db        416 AACCAATGCTGCCCCCAGTGC 436

RESULT 9
US-10-160-719-13
; Sequence 13, Application US/10160719
; Publication No. US20030167028A1
; GENERAL INFORMATION:
;  APPLICANT: Kanwarpal S. Dhugga
;  APPLICANT:  Timothy G. Helentjaris
;  APPLICANT:  Benjamin A. Bowen
;  APPLICANT:  Xun Wang
;  TITLE OF INVENTION: Maize Cellulose Synthases and Uses
;  TITLE OF INVENTION:  Thereof
;  FILE REFERENCE: 0864
;  CURRENT APPLICATION NUMBER: US/10/160,719
;  CURRENT FILING DATE:  2002-06-03
;  PRIOR APPLICATION NUMBER: US/09/371,383A
;  PRIOR FILING DATE: 1999-08-06
;  PRIOR APPLICATION NUMBER: 60/096,822
;  PRIOR FILING DATE: 1998-08-17
;  NUMBER OF SEQ ID NOS: 60
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13
;   LENGTH: 3725
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (179)...(3400)
US-10-160-719-13

Alignment Scores:
Pred. No.:                7.04e-13        Length:          3725
Score:                   161.00          Matches:           23
Percent Similarity:      76.60 %         Conservative:      13
Best Local Similarity:   48.94%          Mismatches:        11
Query Match:             49.69 %         Indels:             0
DB:                      12              Gaps:               0

US-09-980-043-4 (1-56) x US-10-160-719-13 (1-3725)

Qy          8 IleCysHisThrCysGlyGluGlnValGlyHisAspAlaAsnGlyGluLeuPheValAla 27
              :::|||      |:|||:::   |||:::      |||   :||::::::|||||||||
Db        296 GTCTGCCAGATTTGTGGCGACACTGTTGGCGTTTCAGCCACTGGTGATGTCTTTGTTGCC 355

Qy         28 CysHisGluCysSerTyrProMetCysLysSerCysPheGluPheGluIleLysGluGly 47
              |||:::  |||::: :::|||:::|||:::   :|:::|||:::|||   |||||||||
Db        356 TGCAATGAGTGTGCCTTCCCTGTCTGCCGCCCTTGCTATGAGTACGAGCGCAAGGAAGGG 415

Qy         48 ArgLysValCysLeuArgCys 54
                 :::   ||   :::|||
Db        416 AACCAATGCTGCCCCCAGTGC 436

RESULT 10
US-10-160-719-21
; Sequence 21, Application US/10160719
; Publication No. US20030167028A1
; GENERAL INFORMATION:
;  APPLICANT: Kanwarpal S. Dhugga
;  APPLICANT:  Timothy G. Helentjaris
;  APPLICANT:  Benjamin A.  Bowen
;  APPLICANT:  Xun Wang
;  TITLE OF INVENTION: Maize Cellulose Synthases and Uses
;  TITLE OF INVENTION:  Thereof
;  FILE REFERENCE: 0864
;  CURRENT APPLICATION NUMBER: US/10/160,719
;  CURRENT FILING DATE:  2002-06-03
;  PRIOR APPLICATION NUMBER: US/09/371,383A
;  PRIOR FILING DATE: 1999-08-06
;  PRIOR APPLICATION NUMBER: US 60/096,822
;  PRIOR FILING DATE: 1998-08-17
;  NUMBER OF SEQ ID NOS: 60
```

```
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 21
;   LENGTH: 3725
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (179)...(3400)
US-10-160-719-21

Alignment Scores:
Pred. No.:              7.04e-13        Length:         3725
Score:                  161.00          Matches:        37
Percent Similarity:     76.60%          Conservative:   10
Best Local Similarity:  48.94%          Mismatches:     11
Query Match:            49.69%          Indels:         6
DB:                     12              Gaps:           1

US-09-980-043-4 (1-56) x US-10-160-719-21 (1-3725)

Qy           8 IleCysHisThrCysGlyGluGlnValGlyHisAspAlaAsnGlyGluLeuPheValAla 27
               :::|||       ||||:::   ||||'        |||   |||:::::::||||'||||
Db         296 GTCTGCCAGATTTGTGGCGACACTGTTGGCGTTTCAGCCACTGGTGATGTCTTTGTTGCC 355

Qy          28 CysHisGluCysSerTyrProMetCysLysSerCysPheGlu PheGluIleLysGluGly 47
               |||:::|||||:::::|||:::|||:::    :|||:::|||:::|||   |||||||||
Db         356 TGCAATGAGTGTGCCTTCCCTGTCTGCCGCCCTTGCTATGAGTACGAGCGCAAGGAAGGG 415

Qy          48 ArgLysValCysLeuArgCys 54
                  :::   |||   :::|||
Db         416 AACCAATGCTGCCCTCAGTGC 436

RESULT 11
US-10-160-719-41
; Sequence 41, Application US/10160719
; Publication No. US20030167528A1
; GENERAL INFORMATION:
;  APPLICANT: Kanwarpal S. Dhugga
;  APPLICANT:  Timothy G. Helentjaris
;  APPLICANT:  Benjamin A.  Bowen
;  APPLICANT:  Xun Wang
;  TITLE OF INVENTION: Maize Cellulose Synthases and Uses
;  TITLE OF INVENTION:  Thereof
;  FILE REFERENCE: 0864
;  CURRENT APPLICATION NUMBER: US/10/209,059
;  CURRENT FILING DATE:  2002-08-03
;  PRIOR APPLICATION NUMBER: US/09/371,383
;  PRIOR FILING DATE: 1999-08-06
;  PRIOR APPLICATION NUMBER: US 60/096,822
;  PRIOR FILING DATE: 1998-08-17
;  NUMBER OF SEQ ID NOS: 60
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 41
;   LENGTH: 3725
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (179)...(3400)
US-10-160-719-41

Alignment Scores:
Pred. No.:              7.04e-13        Length:         3725
Score:                  161.00          Matches:        37
Percent Similarity:     76.60%          Conservative:   10
Best Local Similarity:  48.94%          Mismatches:     11
Query Match:            49.69%          Indels:         6
DB:                     12              Gaps:           1

US-09-980-043-4 (1-56) x US-10-160-719-41 (1-3725)

Qy           8 IleCysHisThrCysGlyGluGlnValGlyHisAspAlaAsnGlyGluLeuPheValAla 27
               :::|||       ||||:::   ||||'        |||   |||:::::::||||'||||
Db         296 GTCTGCCAGATTTGTGGCGACACTGTTGGCGTTTCAGCCACTGGTGATGTCTTTGTTGCC 355

Qy          28 CysHisGluCysSerTyrProMetCysLysSerCysPheGluPheGluIleLysGluGly 47
               |||:::|||||:::::|||:::|||:::    :|||:::|||:::|||   |||||||||
Db         356 TGCAATGAGTGTGCCTTCCCTGTCTGCCGCCCTTGCTATGAGTACGAGCGCAAGGAAGGG 415

Qy          48 ArgLysValCysLeuArgCys 54
                  :::   |||   :::|||
Db         416 AACCAATGCTGCCCTCAGTGC 436

RESULT 12
US-10-209-059-41
; Sequence 41, Application US/10209059
; Publication No. US20030163838A1
; GENERAL INFORMATION:
;  APPLICANT: Dhugga, Kanwarpal S.
;  APPLICANT:  Wang, Haiyin
;  TITLE OF INVENTION: Maize Cellulose Synthases and Uses
;  TITLE OF INVENTION:  Thereof
;  FILE REFERENCE: 0864R2
;  CURRENT APPLICATION NUMBER: US/10/209,059
;  CURRENT FILING DATE:  2002-07-31
;  PRIOR APPLICATION NUMBER: 60/096,822
;  PRIOR FILING DATE: 1998-08-17
;  PRIOR APPLICATION NUMBER: 09/371,383
;  PRIOR FILING DATE: 1999-08-06
;  PRIOR APPLICATION NUMBER: 09/550,483
;  PRIOR FILING DATE: 2000-04-14
;  NUMBER OF SEQ ID NOS: 52
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 41
;   LENGTH: 3568
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 3487
;   OTHER INFORMATION: n = A,T,C or G
US-10-209-059-41

Alignment Scores:
Pred. No.:              1.81e-12        Length:         3568
Score:                  158.00          Matches:        23
Percent Similarity:     70.37%          Conservative:   15
Best Local Similarity:  42.59%          Mismatches:     16
Query Match:            48.77%          Indels:         0
DB:                     12              Gaps:           0

US-09-980-043-4 (1-56) x US-10-209-059-41 (1-3568)

Qy           1 MetMetGluSerGlyAlaProIleCysHisThrCysGlyGluGlnValGlyHisAspAla 20
               |||   :::             :::|||      ||||||:::   ||||||   :::
Db          63 ATGGACCAGCGGAACGGCCAGGTGTGCCAGATTTGCGGCGACGACGTGGGGCGCAACCCC 122

Qy          21 AsnGlyGluLeuPheValAlaCysHisGluCysSerTyrProMetCysLysSerCysPhe 40
               :::||||||   ||||||||||||:::||||||:::::::|||:::|||:::   |||:::
Db         123 GACGGGGAGCCTTTCGTGGCCTGCAACGAGTGCGCCTTCCCCATCTGCCGGGACTGCTAC 182

Qy          41 GluPheGluIleLysGluGlyArgLysValCysLeuArgCys 54
               |||:::|||   :::||||||   :::   |||   :::|||
Db         183 GAGTACGAGCGCCGCGAGGGCACGCAGAACTGCCCCCAGTGC 224

RESULT 13
US-10-160-719-1
; Sequence 1, Application US/10160719
; Publication No. US20030167528A1
; GENERAL INFORMATION:
;  APPLICANT: Kanwarpal S. Dhugga
;  APPLICANT:  Timothy G. Helentjaris
;  APPLICANT:  Benjamin A.  Bowen
;  APPLICANT:  Xun Wang
;  TITLE OF INVENTION: Maize Cellulose Synthases and Uses
;  TITLE OF INVENTION:  Thereof
```

FILE REFERENCE: 0864
CURRENT APPLICATION NUMBER: US/10/160,719
CURRENT FILING DATE: 2002-06-03
PRIOR APPLICATION NUMBER: US/09/371,383A
PRIOR FILING DATE: 1999-08-06
PRIOR APPLICATION NUMBER: US 60/096,822
PRIOR FILING DATE: 1998-08-17
NUMBER OF SEQ ID NOS: 60
SOFTWARE: FastSEQ for Windows Version 3.0
SEQ ID NO 1
LENGTH: 3568
TYPE: DNA
ORGANISM: Zea mays
FEATURE:
NAME/KEY: CDS
LOCATION: (63)...(3239)
FEATURE:
NAME/KEY: unsure
LOCATION: (3487)...(3487)
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(3568)
OTHER INFORMATION: n = A,T,C or G
S-10-160-719-1

lignment Scores:
red. No.: 1.81e-12 Length: 3568
core: 158.00 Matches: 23
ercent Similarity: 70.37% Conservative: 15
est Local Similarity: 42.59% Mismatches: 16
uery Match: 48.77% Indels: 0
B: 12 Gaps: 0

S-09-980-043-4 (1-56) x US-10-160-719-1 (1-3568)

```
y          1 MetMetGluSerGlyAlaProIleCysHisThrCysGlyGluGlnValGlyHisAspAla 20
             |||   :::          :::|||     ||||||:::  ||||||   :::
b         63 ATGGACCAGCGGAACGGCCAGGTGTGCCAGATTTGCGGCGACGACGTGGGGCGCAACCCC 122

y         21 AsnGlyGluLeuPheValAlaCysHisGluCysSerTyrProMetCysLysSerCysPhe 40
             :::|||||   ||||||||||||:::||||||:::::::||:::|||:::   ||:::
b        123 GACGGGGAGCCTTTCGTGGCCTGCAACGAGTGCGCCTTCCCCATCTGCCGGGACTGCTAC 182

y         41 GluPheGluIleLysGluGlyArgLysValCysLeuArgCys 54
             |||:::|||   :::||||||   :::   |||   :::|||
b        183 GAGTACGAGCGCCGCGAGGGCACGCAGAACTGCCCCCAGTGC 224
```

ESULT 14
S-09-900-237-3
Sequence 3, Application US/09900237
Patent No. US20020120124A1
GENERAL INFORMATION:
APPLICANT: Allen, Stephen
TITLE OF INVENTION: Plant Cellulose Synthases
FILE REFERENCE: BB1170 US CIP
CURRENT APPLICATION NUMBER: US/09/900,237
CURRENT FILING DATE: 2001-07-06
PRIOR APPLICATION NUMBER: 60/092,844
PRIOR FILING DATE: 1998-07-14
PRIOR APPLICATION NUMBER: PCT/US99/15871
PRIOR FILING DATE: 1999-07-13
PRIOR APPLICATION NUMBER: 09/720383
PRIOR FILING DATE: 2000-12-21
NUMBER OF SEQ ID NOS: 33
SOFTWARE: Microsoft Office 97
SEQ ID NO 3
LENGTH: 3776
TYPE: DNA
ORGANISM: Zea mays
S-09-900-237-3

lignment Scores:
red. No.: 1.94e-12 Length: 3776

Score: 158.0 Matches: 23
Percent Similarity: 70.3 % Conservative: 15
Best Local Similarity: 42.6 % Mismatches: 16
Query Match: 48.7 % Indels: 0
DB: 10 Gaps: 0

US-09-980-043-4 (1-56) x US-09-900-237-3 (1-3776)

```
Qy          1 MetMetGluSerGl AlaProIleCysH.sThrCysGlyGluGlnValGlyHisAspAla 20
              |||   :::          :::|||     ||||||:::  ||||||   :::
Db        269 ATGGACCAGCGGAA CGGCAGGTGTGCCAGATTTGCGGCGACGACGTGGGGCGCAACCCC 328

Qy         21 AsnGlyGluLeuPhe ValAlaCysHisGluCysSerTyrProMetCysLysSerCysPhe 40
              :::|||||   ||||||||||||:::||||||:::::::||:::|||:::   ||:::
Db        329 GACGGGGAGCCTTT TCGTGGCCTGCAACGAGTGCGCCTTCCCCATCTGCCGGGACTGCTAC 388

Qy         41 GluPheGluIleLy GluGlyArgLysValCysLeuArgCys 54
              |||:::|||   :: ||||||   :::   |||   :::|||
Db        389 GAGTACGAGCGCCG CGAGGGCACGCAGAACTGCCCCCAGTGC 430
```

RESULT 15
US-09-900-237-13
; Sequence 13, Application US/09900237
; Patent No. US20020120124A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Stephen
; TITLE OF INVENTION: Plant Cellulose Synthases
; FILE REFERENCE: BB1170 US CIP
; CURRENT APPLICATION NUMBER: US/09/900,237
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: 60/092,844
; PRIOR FILING DATE: 1998-07-14
; PRIOR APPLICATION NUMBER: PCT/US99/15871
; PRIOR FILING DATE: 1999-07-13
; PRIOR APPLICATION NUMBER: 09/720383
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 13
; LENGTH: 3517
; TYPE: DNA
; ORGANISM: Glycine max
US-09-900-237-13

Alignment Scores:
Pred. No.: 6.7e-12 Length: 3517
Score: 154.0 Matches: 22
Percent Similarity: 74.4 % Conservative: 13
Best Local Similarity: 46.8 % Mismatches: 12
Query Match: 47.5 % Indels: 0
DB: 10 Gaps: 0

US-09-980-043-4 (1-56) x US-09-900-237-13 (1-3517)

```
Qy          8 IleCysHisThrCy GlyGluGlnValGlyHisAspAlaAsnGlyGluLeuPheValAla 27
              :::|||     || |||:::   |: |||        :::||:.:||||  ||||
Db        263 GTGTGTGAGATTTG GGTGATGGCGTGGGACT ACGGTGGATGGAGACTTGTTTGTGGCT 322

Qy         28 CysHisGluCysSe TyrProMetCysLysSerCysPheGluPheGluIleLysGluGly 47
              :||:::|||.:|   :::|:.:::: |:::   ||:::||.:::|||   :::||||:
Db        323 TGCAATGAGTGTG TTTCCAGTGTGCAGGCCTTGCTATGAGTATGAAAGGAGAGAAGGA 382

Qy         48 ArgLysValCysLe ArgCys 54
                    :::|||   :::|||
Db        383 AGCCACCTTTGCCC CAGTGC 403
```

Search completed: October 31, 2003, 23:44:25
Job time : 44.7853 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 31, 2003, 09:06:20 ; Search time 8.88281 Seconds
(without alignments)
2782.867 Million cell updates/sec

Title: US-09-980-043-4
Perfect score: 324
Sequence: 1 MMESGAPICHTCGEQVGHTA.....KSCPEPETKEGREVCLRCGS 56

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=x1h
Q=/cgn2_1/USPTO_spool/US09980043/runat_29102003_163853_21883/app_query.fasta_1.1358
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09980043_aCGN_1_1_76_@runat_29102003_163853_21883 -NCPU=6 -ICPU=3
-NO_MMAP LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 LONGLOG
DEV_TIMEOUT=120 WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	264	81.5	3328	3	US-08-960-048-1	Sequence 1, Appli
2	264	81.5	3328	4	US-09-838-586-1	Sequence 1, Appli
3	161	49.7	1741	4	US-09-221-013A-13	Sequence 13, Appl
4	152	46.9	3603	4	US-09-221-013A-5	Sequence 5, Appli
5	152	46.9	3673	4	US-09-221-013A-11	Sequence 11, Appl
6	152	46.9	8411	4	US-09-221-013A-3	Sequence 3, Appli
7	148	45.7	3828	4	US-09-221-013A-7	Sequence 7, Appli
8	148	45.7	5009	4	US-09-221-013A-4	Sequence 4, Appli
9	138	42.6	3614	4	US-09-221-013A-9	Sequence 9, Appli
10	87	26.9	1063	3	US-08-960-048-3	Sequence 3, Appli
11	87	26.9	1063	4	US-09-838-586-3	Sequence 3, Appli
12	77.5	23.9	1397	3	US-08-946-026-17	Sequence 17, Appl
13	71.5	22.1	9319	4	US-08-976-259-85	Sequence 85, Appl
c 14	69.5	21.5	2073	4	US-09-221-017B-894	Sequence 894, App
c 15	68	21.0	861	4	US-09-107-532A-1732	Sequence 1732, Ap
16	67	20.7	1805	4	US-09-016-434-1086	Sequence 1086, Ap
17	67	20.7	4201	1	US-08-080-255-4	Sequence 4, Appli
18	67	20.7	4201	3	US-08-465-713-4	Sequence 4, Appli
19	67	20.7	4201	5	PCT-US93-05857-4	Sequence 4, Appli
20	67	20.7	11907	3	US-08-061-376-4	Sequence 4, Appli
21	67	20.7	14255	1	US-08-320-559-1	Sequence 1, Appli
22	67	20.7	14255	1	US-08-327-392-1	Sequence 1, Appli
23	67	20.7	14255	1	US-08-306-691B-55	Sequence 55, Appl
24	67	20.7	14255	3	US-08-545-860D-1	Sequence 1, Appli
25	67	20.7	14255	5	PCT-US94-04496-1	Sequence 1, Appli
26	66	20.4	246	4	US-09-016-434-442	Sequence 442, App
27	66	20.4	3262	2	US-08-678-039A-41	Sequence 41, Appl
28	65.5	20.2	374	3	US-08-946-026-42	Sequence 42, Appl
29	65.5	20.2	571	3	US-09-328-111-620	Sequence 620, App
30	65.5	20.2	612	4	US-09-252-991A-16101	Sequence 16101, A
31	65.5	20.2	1234	4	US-09-166-350-24	Sequence 24, Appl
32	65.5	20.2	1246	1	US-08-269-441A-1	Sequence 1, Appli
33	65.5	20.2	1246	3	US-09-276-851-1	Sequence 1, Appli
34	65.5	20.2	2220	4	US-09-252-991A-16207	Sequence 16207, A
c 35	65.5	20.2	3744	4	US-09-252-991A-16389	Sequence 16389, A
36	65	20.1	1824	3	US-09-030-335-1	Sequence 1, Appli
37	65	20.1	2098	2	US-08-937-931-5	Sequence 5, Appli
38	65	20.1	2098	3	US-09-285-502-5	Sequence 5, Appli
39	65	20.1	2098	4	US-09-709-126-5	Sequence 5, Appli
40	65	20.1	2098	4	US-09-871-385A-5	Sequence 5, Appli
41	65	20.1	2625	3	US-09-030-335-8	Sequence 8, Appli
42	65	20.1	2763	3	US-09-030-335-3	Sequence 3, Appli
43	65	20.1	2796	2	US-08-937-931-3	Sequence 3, Appli
44	65	20.1	2796	3	US-09-285-502-3	Sequence 3, Appli
45	65	20.1	2796	4	US-09-709-126-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-08-960-048-1
; Sequence 1, Application US/08960048C
; Patent No. 6271443
; GENERAL INFORMATION:
; APPLICANT: Stalker, D. et al.
; TITLE OF INVENTION: Plant Cellulose Synthase and Promoter
; TITLE OF INVENTION: Sequences
; FILE REFERENCE: 15621/01/US
; CURRENT APPLICATION NUMBER: US/08/960,048C
; CURRENT FILING DATE: 1997-10-29
; PRIOR APPLICATION NUMBER: 60/029,987
; PRIOR FILING DATE: 1996-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 3328
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Oligonucleotide
US-08-960-048-1

Alignment Scores:

Pred. No.:	3.39e-25	Length:	3328
Score:	264.00	Matches:	42
Percent Similarity:	89.29%	Conservative:	8
Best Local Similarity:	75.00%	Mismatches:	6
Query Match:	81.48%	Indels:	0
DB:	3	Gaps:	0

US-09-980-043-4 (1-56) x US-08-960-048-1 (1-3328)

```
Qy          1 MetMetGluSerGlyAlaProIleCysHisThrCysGlyGluGlnValGlyHisAspAla 20
              ||| |||||||||||   | |:::||||||||||||||||||||   ||||||   :::
Db        179 ATGATGGAATCTGGGGTTCCTGTTTGCCACACTTGTGGTGAACATGTTGGGTTGAATGTT 238
```

```
y          21 AsnGlyGluLeuPheValAlaCysHisGluCysSerTyrProMetCysLysSerCysPhe 40
              |||||||||   |||||||||||||||||||||||:::::|||:::|||||||||||||||
o         239 AATGGTGAACCTTTTGTGGCTTGCCATGAATGTAATTTCCCTATTTGTAAGAGTTGTTTT 298

y          41 GluPheGluIleLysGluGlyArgLysValCysLeuArgCysGlySer 56
              |||:::::::::|||||||||||||||   |||||||||||||||||
o         299 GAGTATGATCTTAAGGAAGGACGAAAAGCTTGCTTGCGTTGTGGTAGT 346

ESULT 2
S-09-838-586-1
 Sequence 1, Application US/09838586
 Patent No. 6576818
 GENERAL INFORMATION:
  APPLICANT: Stalker, D. et al.
  TITLE OF INVENTION: Plant Cellulose Synthase and Promoter
  TITLE OF INVENTION:  Sequences
  FILE REFERENCE: 15621/02/US
  CURRENT APPLICATION NUMBER: US/09/838,586
  CURRENT FILING DATE:  2001-04-18
  PRIOR APPLICATION NUMBER: 60/029,987
  PRIOR FILING DATE: 1996-10-29
  PRIOR APPLICATION NUMBER: 08/960,048
  PRIOR FILING DATE: 1997-10-29
  NUMBER OF SEQ ID NOS: 12
  SOFTWARE: FastSEQ for Windows Version 4.0
 SEQ ID NO 1
   LENGTH: 3328
   TYPE: DNA
   ORGANISM: Artificial Sequence
   FEATURE:
   OTHER INFORMATION: Synthetic Oligonucleotide
S-09-838-586-1

lignment Scores:
red. No.:                3.39e-25        Length:         3328
core:                    264.00          Matches:        42
ercent Similarity:       89.29%          Conservative:   8
est Local Similarity:    75.00%          Mismatches:     6
uery Match:              81.48%          Indels:         0
B:                       4               Gaps:           0

S-09-980-043-4 (1-56) x US-09-838-586-1 (1-3328)

y           1 MetMetGluSerGlyAlaProIleCysHisThrCysGlyGluGlnValGlyHisAspAla 20
              |||||||||||||||   |||:::|||||||||||||||||||   ||||||   :::
b         179 ATGATGGAATCTGGGGTTCCTGTTTGCCACACTTGTGGTGAACATGTTGGGTTGAATGTT 238

y          21 AsnGlyGluLeuPheValAlaCysHisGluCysSerTyrProMetCysLysSerCysPhe 40
              |||||||||   |||||||||||||||||||||||:::::|||:::|||||||||||||||
b         239 AATGGTGAACCTTTTGTGGCTTGCCATGAATGTAATTTCCCTATTTGTAAGAGTTGTTTT 298

y          41 GluPheGluIleLysGluGlyArgLysValCysLeuArgCysGlySer 56
              |||:::::::::|||||||||||||||   |||||||||||||||||
b         299 GAGTATGATCTTAAGGAAGGACGAAAAGCTTGCTTGCGTTGTGGTAGT 346

ESULT 3
S-09-221-013A-13
 Sequence 13, Application US/09221013A
 Patent No. 6495740
 GENERAL INFORMATION:
  APPLICANT:  Arioli, Antonio
  APPLICANT:  Williamson, Richard E.
  APPLICANT:  Betzner, Andreas S.
  APPLICANT:  Peng, Liangcai
  TITLE OF INVENTION:  Manipulation of cellulose and/or Beta-1,4-glucan
  FILE REFERENCE:  96-98
  CURRENT APPLICATION NUMBER: US/09/221,013A
  CURRENT FILING DATE:  1998-12-23
  PRIOR APPLICATION NUMBER:  PCT/AU97/00402
  PRIOR FILING DATE:  1997-06-24
  PRIOR APPLICATION NUMBER:  AU PO0699
  PRIOR FILING DATE:  1996-06-27
;  NUMBER OF SEQ ID NOS:  37
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 13
;   LENGTH: 1741
;   TYPE: DNA
;   ORGANISM: Oryza sativa
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (101)..(1741)
US-09-221-013A-13

Alignment Scores:
Pred. No.:                5.7[illegible]-12       Length:         1741
Score:                    161.[illegible]         Matches:        23
Percent Similarity:       76.[illegible]%         Conservative:   13
Best Local Similarity:    48.[illegible]%         Mismatches:     11
Query Match:              49.[illegible]%         Indels:         0
DB:                       4                       Gaps:           0

US-09-980-043-4 (1-56) x US-09-221-013A-13 (101-1741)

Qy          8 IleCysHisThrCysGlyGluGlnValGlyHisAspAlaAsnGlyGluLeuPheValAla 27
              :::|||           |  :::   ,:|          |||   |||:::::::.||||:
Db        218 GTCTGCCACGATTC[illegible]GGCGACACTGTTGGCCACTGGCGACGTCTTTGTTGCC 277

Qy         28 CysHisGluCysSerTyrProMetCysLysSerCysPheGluPheGluIleLysGluGly 47
              |||:::|||||:::::::|||:::|||:::   | ,:::|||:::||    ||'|||||
Db        278 TGCAATGAGTGCGG[illegible]TTCCCGGTTTGCCGCCCTTGCTACGAGTACGAACGCAAGGAAGGG 337

Qy         48 ArgLysValCysLeuArgCys 54
                 :::  |||  :::|'|
Db        338 AACCAGTGCTGCCC[illegible]AGTGC 358

RESULT 4
US-09-221-013A-5
; Sequence 5, Application US/09221013A
; Patent No. 6495740
; GENERAL INFORMATION:
;  APPLICANT:  Arioli, Antonio
;  APPLICANT:  Williamson, Richard E.
;  APPLICANT:  Betzner, Andreas S.
;  APPLICANT:  Peng, Liangca.
;  TITLE OF INVENTION:  Manipulation of cellulose and/or Beta-1,4 glucan
;  FILE REFERENCE:  96-98
;  CURRENT APPLICATION NUMBER: US/09/221,013A
;  CURRENT FILING DATE:  1998-12-23
;  PRIOR APPLICATION NUMBER:  PCT/AU97/00402
;  PRIOR FILING DATE:  1997-06-24
;  PRIOR APPLICATION NUMBER:  AU PO0699
;  PRIOR FILING DATE:  1996-06-27
;  NUMBER OF SEQ ID NOS:  37
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 5
;   LENGTH: 3603
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(3243)
US-09-221-013A-5

Alignment Scores:
Pred. No.:                2.29e-10        Length:         3603
Score:                    152.00          Matches:        22
Percent Similarity:       74.47%          Conservative:   13
Best Local Similarity:    46.81%          Mismatches:     12
Query Match:              46.91%          Indels:         0
DB:                       4               Gaps:           0

US-09-980-043-4 (1-56) x US-09-221-013A-5 (1-3603)

Qy          8 IleCysHisThrCysGlyGluGlnValGlyHisAspAlaAsnGlyGluLeuPheValAla 27
              ||||||      || |||:::   ||:|||            |||:::::|||:|:||||
```

```
Db        112 ATATGTCAGATCTGTGGTGATGATGTTGGACTCGCTGAAACTGGAGATGTCTTTGTCGCG 171

Qy         28 CysHisGluCysSerTyrProMetCysLysSerCysPheGluPheGluIleLysGluGly 47
              |||:::|||||:::::::|||:::||:::   |||:::|: :::|:|   |||:::|||
Db        172 TGTAATGAATGTGCCTTCCCTGTGTGTCGGCCTTGCTATGA:TACGAGAGGAAAGATGGA 231

Qy         48 ArgLysValCysLeuArgCys 54
                 :::   |||   :::|||
Db        232 ACTCAGTGTTGCCCTCAATGC 252

RESULT 5
US-09-221-013A-11
; Sequence 11, Application US/09221013A
; Patent No. 6495740
; GENERAL INFORMATION:
;  APPLICANT:  Arioli, Antonio
;  APPLICANT:  Williamson, Richard E.
;  APPLICANT:  Betzner, Andreas S.
;  APPLICANT:  Peng, Liangcai
;  TITLE OF INVENTION:  Manipulation of cellulose and/or Beta 1,4-glucan
;  FILE REFERENCE:  96-98
;  CURRENT APPLICATION NUMBER: US/09/221,013A
;  CURRENT FILING DATE:  1998-12-23
;  PRIOR APPLICATION NUMBER:  PCT/AU97/00402
;  PRIOR FILING DATE:  1997-06-24
;  PRIOR APPLICATION NUMBER:  AU PO0699
;  PRIOR FILING DATE:  1996-06-27
;  NUMBER OF SEQ ID NOS:  37
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 11
;   LENGTH: 3673
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (71)..(3313)
US-09-221-013A-11

Alignment Scores:
Pred. No.:                2.35e-10        Length:         3673
Score:                    152.00          Matches:        22
Percent Similarity:       74.47%          Conservative:   13
Best Local Similarity:    46.81%          Mismatches:     12
Query Match:              46.91%          Indels:         0
DB:                       4               Gaps:           0

US-09-980-043-4 (1-56) x US-09-221-013A-11 (71-3313)

Qy           6 IleCysHisThrCysGlyGluGlnValGlyHisAspAlaAsnGlyGluLeuPheValAla 27
               |||:|      |||||:::  ||||||           |||:::::|||||||||
Db         181 ATATGTCAGATCTGTGGTGATGATGTTGGACTCGCTGAAACTGGAGATGTCTTTGTCGCG 240

Qy          28 CysHisGluCysSerTyrProMetCysLysSerCysPheGluPheGluIleLysGluGly 47
               |||:::|||||:::::::|||:::||:::   |||:::|: :::|:|   |||:::|||
Db         241 TGTAATGAATGTGCCTTCCCTGTGTGTCGGCCTTGCTATGA:TACGAGAGGAAAGATGGA 300

Qy          48 ArgLysValCysLeuArgCys 54
                 :::   |||   :::|||
Db         301 ACTCAGTGTTGCCCTCAATGC 321

RESULT 6
US-09-221-013A-3
; Sequence 3, Application US/09221013A
; Patent No. 6495740
; GENERAL INFORMATION:
;  APPLICANT:  Arioli, Antonio
;  APPLICANT:  Williamson, Richard E.
;  APPLICANT:  Betzner, Andreas S.
;  APPLICANT:  Peng, Liangcai
;  TITLE OF INVENTION:  Manipulation of cellulose and/or Beta 1,4-glucan
;  FILE REFERENCE:  96-98
;  CURRENT APPLICATION NUMBER: US/09/221,013A
;  CURRENT FILING DATE:  1998-12-23
;  PRIOR APPLICATION NUMBER:  PCT/AU97/00402
;  PRIOR FILING DATE:  1997-06-24
;  PRIOR APPLICATION NUMBER:  AU PO0699
;  PRIOR FILING DATE:  1996-06-27
;  NUMBER OF SEQ ID NOS:  37
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 8411
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
US-09-221-013A-3

Alignment Scores:
Pred. No.:                6.9e-10         Length:         8411
Score:                    152.00          Matches:        22
Percent Similarity:       74.47%          Conservative:   13
Best Local Similarity:    46.81%          Mismatches:     12
Query Match:              46.91%          Indels:         0
DB:                       4               Gaps:           0

US-09-980-043-4 (1-56) x US-09-221-013A-3 (1-8411)

Qy           8 IleCysHisThrCysGlyGluGlnValGlyHisAspAlaAsnGlyGluLeuPheValAla 27
               ||||||      |||||:::  ||||||           |||:::::|||||||||
Db        2934 ATATGTCAGATCTGTGGTGATGATGTTGGACTCGCTGAAACTGGAGATGTCTTTGTCGCG 2993

Qy          28 CysHisGluCysSerTyrProMetCysLysSerCysPheGluPheGluIleLysGluGly 47
               |||:::|||||:::::::|||:::||:::   |||:::|||:::|||   |||:::|||
Db        2994 TGTAATGAATGTGCCTTCCCTGTGTGTCGGCCTTGCTATGAGTACGAGAGGAAAGATGGA 3053

Qy          48 ArgLysValCysLeuArgCys 54
                 :::   |||   :::|||
Db        3054 ACTCAGTGTTGCCCTCAATGC 3074

RESULT 7
US-09-221-013A-7
; Sequence 7, Application US/09221013A
; Patent No. 6495740
; GENERAL INFORMATION:
;  APPLICANT:  Arioli, Antonio
;  APPLICANT:  Williamson, Richard E.
;  APPLICANT:  Betzner, Andreas S.
;  APPLICANT:  Peng, Liangcai
;  TITLE OF INVENTION:  Manipulation of cellulose and/or Beta-1,4-glucan
;  FILE REFERENCE:  96-98
;  CURRENT APPLICATION NUMBER: US/09/221,013A
;  CURRENT FILING DATE:  1998-12-23
;  PRIOR APPLICATION NUMBER:  PCT/AU97/00402
;  PRIOR FILING DATE:  1997-06-24
;  PRIOR APPLICATION NUMBER:  AU PO0699
;  PRIOR FILING DATE:  1996-06-27
;  NUMBER OF SEQ ID NOS:  37
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 7
;   LENGTH: 3828
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (239)..(3490)
US-09-221-013A-7

Alignment Scores:
Pred. No.:                8.36e-10        Length:         3828
Score:                    148.00          Matches:        21
Percent Similarity:       66.67%          Conservative:   15
Best Local Similarity:    38.89%          Mismatches:     18
Query Match:              45.68%          Indels:         0
DB:                       4               Gaps:           0

US-09-980-043-4 (1-56) x US-09-221-013A-7 (1-3828)

Qy           1 MetMetGlnSerGlyAlaProIleCysHisThrCysGlyGluGlnValGlyHisAspAla 20
```

```
Db        329 GTACAAGAACTGAGTGGGCAAACATGTCAAATCTGTGGAGATGAAATCGAATTAACGGTT 388

Qy         21 AsnGlyGluLeuPheValAlaCysHisGluCysSerTyrProMetCysLysSerCysPhe 40
              :::  |||||||||||||||||||:::||||||:::::|||:::|||:::   |||:::
Db        389 AGCAGTGAGCTCTTTGTTGCTTGCAACGAATGCGCATTCCCGGTTTGTAGACCATGCTAT 448

Qy         41 GluPheGluIleLysGluGlyArgLysValCysLeuArgCys 54
              |||:::|||   :::||||||   :::   |||   :::|||
Db        449 GAGTATGAACGTAGAGAAGGAAATCAAGCTTGTCCTCAGTGC 490

ESULT 8
S-09-221-013A-4
 Sequence 4, Application US/09221013A
 Patent No. 6495740
 GENERAL INFORMATION:
  APPLICANT:  Arioli, Antonio
  APPLICANT:  Williamson, Richard E.
  APPLICANT:  Betzner, Andreas S.
  APPLICANT:  Peng, Liangcai
  TITLE OF INVENTION:  Manipulation of cellulose and/or Beta-1,4-glucan
  FILE REFERENCE:  96-98
  CURRENT APPLICATION NUMBER: US/09/221,013A
  CURRENT FILING DATE:  1998-12-23
  PRIOR APPLICATION NUMBER:  PCT/AU97/00402
  PRIOR FILING DATE:  1997-06-24
  PRIOR APPLICATION NUMBER:  AU PO0699
  PRIOR FILING DATE:  1996-06-27
  NUMBER OF SEQ ID NOS:  37
  SOFTWARE:  PatentIn Ver. 2.0
 SEQ ID NO 4
   LENGTH: 5009
   TYPE: DNA
   ORGANISM: Arabidopsis thaliana
   FEATURE:
   NAME/KEY: misc_feature
   LOCATION: (1)..(5009)
   OTHER INFORMATION: N is A,  T,  G  or C.
S-09-221-013A-4

lignment Scores:
red. No.:                1.19e-09        Length:         5009
core:                    148.00          Matches:        21
ercent Similarity:       66.67%          Conservative:   15
est Local Similarity:    38.89%          Mismatches:     18
uery Match:              45.68%          Indels:         0
B:                       4               Gaps:           0

S-09-980-043-4 (1-56) x US-09-221-013A-4 (1-5009)

Qy          1 MetMetGluSerGlyAlaProIleCysHisThrCysGlyGluGlnValGlyHisAspAla 20
              :::  |||                 |||      |||||::::::::::
Db       1463 GTACAAGAACTGAGTGGGCAAACATGTCAAATCTGTGGAGATGAAATCGAATTAACGGTT 1522

Qy         21 AsnGlyGluLeuPheValAlaCysHisGluCysSerTyrProMetCysLysSerCysPhe 40
              :::  |||||||||||||||||||:::||||||:::::|||:::|||:::   |||:::
Db       1523 AGCAGTGAGCTCTTTGTTGCTTGCAACGAATGCGCATTCCCGGTTTGTAGACCATGCTAT 1582

Qy         41 GluPheGluIleLysGluGlyArgLysValCysLeuArgCys 54
              |||:::|||   :::||||||   :::   |||   :::|||
Db       1583 GAGTATGAACGTAGAGAAGGAAATCAAGCTTGTCCTCAGTGC 1624

ESULT 9
S-09-221-013A-9
 Sequence 9, Application US/09221013A
 Patent No. 6495740
 GENERAL INFORMATION:
  APPLICANT:  Arioli, Antonio
  APPLICANT:  Williamson, Richard E.
  APPLICANT:  Betzner, Andreas S.
  APPLICANT:  Peng, Liangcai
  TITLE OF INVENTION:  Manipulation of cellulose and/or Beta-1,4-glucan
  FILE REFERENCE:  96-98
;  CURRENT APPLICATION NUMBER:  US/09/221,013A
;  CURRENT FILING DATE:  1998-12-23
;  PRIOR APPLICATION NUMBER:  PCT/AU97/00402
;  PRIOR FILING DATE:  1997-06-24
;  PRIOR APPLICATION NUMBER:  AU PO0699
;  PRIOR FILING DATE:  1996-06-27
;  NUMBER OF SEQ ID NOS:  37
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 9
;   LENGTH: 3614
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (217)..(1911)
US-09-221-013A-9

Alignment Scores:
Pred. No.:                1.19e-09        Length:         3614
Score:                    148.00          Matches:        25
Percent Similarity:       66.67%          Conservative:   11
Best Local Similarity:    38.89%          Mismatches:     14
Query Match:              45.68%          Indels:         17
DB:                       4               Gaps:           2

US-09-980-043-4 (1-56) x US-09-221-013A-9 (217-3614)

Qy          2 MetGluSer-------------GlyAlaProIle---------------------Cys 9
              |||||:|||                ||    ||:::                   |
Db        217 ATGGAATCCGAAGTTGGAAACCGCGGCAAAGCCGATGAAGAACATTGTTCCGCAAGCTTGC 276

Qy         10 HisThrCysGlyGluGlnValGlyHisAspAlaAsnGlyGluLeuPheValAlaCysHis 29
              ||     :  ||||:|        :::||:::   : ||||:|||
Db        277 CAGATCTGTAGTGAAATGTTGGCAAGATGTTGATGGAGATCGTTTTGTGGCTTGTGAT 336

Qy         30 GluCysSerTyrProMetCysLysSerCysPheGluPheGluIleLysGluGlyArgLys 49
              |||:||:::||:::||  :::    ||  : ||:::|||   : |:::  :    ::
Db        337 ATTTGTTCAATCCCAGTTTGTCCGGCCTTGTACGAGTATGAGAGGAAAGATTGGAATCAA 396

Qy         50 ValCysLeuArgCys 54
              |:|   :::|||
Db        397 TCTTGTCCTCAGTGC 411

RESULT 10
US-08-960-048-3
; Sequence 3, Application US/08960048C
; Patent No. 6271443
; GENERAL INFORMATION:
;  APPLICANT:  Stalker, D. et al.
;  TITLE OF INVENTION:  Plant Cellulose Synthase and Promoter
;  TITLE OF INVENTION:  Sequences
;  FILE REFERENCE:  15621/01/US
;  CURRENT APPLICATION NUMBER: US/08/960,048C
;  CURRENT FILING DATE:  1997-10-29
;  PRIOR APPLICATION NUMBER:  60/029,987
;  PRIOR FILING DATE:  1996-10-29
;  NUMBER OF SEQ ID NOS:  17
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 1063
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic Oligonucleotide
US-08-960-048-3

Alignment Scores:
Pred. No.:                0.0138          Length:         1063
Score:                    87.00           Matches:        14
Percent Similarity:       88.24%          Conservative:   1
Best Local Similarity:    82.35%          Mismatches:     2
Query Match:              26.65%          Indels:         0
DB:                       3               Gaps:           0
```

```
US-09-980-043-4 (1-56) x US-08-960-048-3 (1-1063)

Qy          1 MetMetGluSerGlyAlaProIleCysHisThrCysGlyGluGlnValGly 17
              |||||||||||||||   |||:::|:|:|||:|||:|||:   |||||
Db        951 ATGATGGAATCTGGGGTTCCTGTTTGCCACACTTGTGGTGAACATGTTGGG 1001

RESULT 11
US-09-838-586-3
; Sequence 3, Application US/09838586
; Patent No. 6576818
; GENERAL INFORMATION:
;  APPLICANT: Stalker, D. et al.
;  TITLE OF INVENTION: Plant Cellulose Synthase and Promoter
;  TITLE OF INVENTION: Sequences
;  FILE REFERENCE: 15621/02/US
;  CURRENT APPLICATION NUMBER: US/09/838,586
;  CURRENT FILING DATE: 2001-04-18
;  PRIOR APPLICATION NUMBER: 60/029,987
;  PRIOR FILING DATE: 1996-10-29
;  PRIOR APPLICATION NUMBER: 08/960,049
;  PRIOR FILING DATE: 1997-10-29
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 1063
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic Oligonucleotide
US-09-838-586-3

Alignment Scores:
Pred. No.:                0.0178          Length:          1063
Score:                    87.00           Matches:           16
Percent Similarity:       88.24%          Conservative:       1
Best Local Similarity:    82.35%          Mismatches:         2
Query Match:              26.85%          Indels:             0
DB:                       4               Gaps:               0

US-09-980-043-4 (1-56) x US-09-838-586-3 (1-1063)

Qy          1 MetMetGluSerGlyAlaProIleCysHisThrCysGlyGlnGlnValGly 17
              ||||:|:|||||||   |||:::|||:|||:|||||||:   |||:|
Db        951 ATGATGGAATCTGGGGTTCCTGTTTGCCACACTTGTGGTGAACATGTTGGG 1001

RESULT 12
US-08-946-026-17
; Sequence 17, Application US/08946026
; Patent No. 6034218
;  GENERAL INFORMATION:
;    APPLICANT:  Reed, Steven G.
;    APPLICANT:  Dillon, Davin C.
;    APPLICANT:  Twardzik, Daniel R.
;    APPLICANT:  Mitcham, Jennifer L.
;    TITLE OF INVENTION:  COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
;    TITLE OF INVENTION:  AND IMMUNODIAGNOSIS OF PROSTATE CANCER
;    NUMBER OF SEQUENCES:  59
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/946,026
;      FILING DATE:  07-OCT-1997
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
;      REGISTRATION NUMBER:  31,392
;      REFERENCE/DOCKET NUMBER:  210121.424C1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  17:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1397 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-946-026-17

Alignment Scores:
Pred. No.:                0.455           Length:          1397
Score:                    77.50           Matches:           18
Percent Similarity:       38.89%          Conservative:      10
Best Local Similarity:    25.00%          Mismatches:        21
Query Match:              23.92%          Indels:            23
DB:                       3               Gaps:               2

US-09-980-043-4 (1-56) x US-08-946-026-17 (1-1397)

Qy          7 ProIleCysHisThrCysGlyGluGlnValGlyHisAspAlaAsnGlyGluLeuPheVal 26
              |||   |||       |||   |||::::::            ||||||   |||
Db        110 CCCAAGTGCACGGTGTGTCTGGAGCGCATGGACGAGTCTGTGAATGGCATCCTCACAACG 169

Qy         27 AlaCysHis----------------------------------GluCysSerTyrPro 34
              |||:::                                     :::   :::   |||
Db        170 TTATGTAACCACATCTTCCACAGCCAGTGTCTACAGCGCTGGGACGATACCACGTGTCCT 229

Qy         35 MetCysLys----------------------------------SerCysPheGluPheGlu 43
              :::|||:::                                 |||||||||
Db        230 GTTTGCCGGTACTGTCAAACGCCCGAGCCAGTAGAAGAAAATAAGTGTTTTGAGTGTGGT 289

Qy         44 IleLysGluGlyArgLysValCysLeuArgCysGly 55
              ::::::|||         :::||||||   ||||||
Db        290 GTTCAGGAAAATCTTTGGATTTGTTTAATATGCGGC 325

RESULT 13
US-08-976-259-85
; Sequence 85, Application US/08976259
; Patent No. 6316609
;  GENERAL INFORMATION:
;    APPLICANT:  Dillon, Patrick J.
;    APPLICANT:  Choi, Gil H.
;    APPLICANT:  Welch, Rodney A.
;    TITLE OF INVENTION:  Nucleotide Sequence of Escherichia coli
; Patent No. 6316609
;    NUMBER OF SEQUENCES:  142
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sterne, Kessler, Goldstein & Fox P.L.L.C.
;      STREET:  1100 New York Ave, N.W., Suite 600
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4Mb storage
;      COMPUTER:  HP Vectra 486/33
;      OPERATING SYSTEM:  MSDOS version 6.2
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/976,259
;      FILING DATE:  Herewith
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/031,626 AND US 60/061,953
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Steffe, Eric K.
```

```
      REGISTRATION NUMBER:  36,688
      REFERENCE/DOCKET NUMBER:  1488.0740002/EKS/CBM
    TELECOMMUNICATION INFORMATION:
      TELEPHONE:  (202) 371-2600
      TELEFAX:  (202) 371-2540
  INFORMATION FOR SEQ ID NO:  85:
    SEQUENCE CHARACTERISTICS:
      LENGTH:  9319 base pairs
      TYPE:  nucleic acid
      STRANDEDNESS:  double
      TOPOLOGY:  linear
US-08 976-259-85

Alignment Scores:
red. No.:                33.4          Length:          9319
core:                    71.50         Matches:         17
ercent Similarity:       47.73%        Conservative:    4
est Local Similarity:    38.64%        Mismatches:      16
Query Match:             22.07%        Indels:          7
DB:                      4             Gaps:            2

US-09-980-043-4 (1-56) x US-08-976-259-85 (1-9319)

Qy          20 AlaAsnGlyGluLeuPheValAlaCysHisGlu-----------CysSerTyrProMet 35
               |||   ||||||||||||         |||              |||
Db        8463 GCTGTGGGTGAGCTTTTCACTGGCGTTCACCACCATGGTCTGTTGTGCTGCCTGCGTACC 8522

Qy          36 CysLysSerCysPhe---------GluPheGluIleLysGluGlyArgLysValCysLeu 52
               |||||||:::|||             :::   ||||||            |||:::|||
Db        8523 TGCAAGACGTGCATCTCTGGCGTTGATGCTGAGATTTTTACCGCTCTGAAGCTGTGCGCC 8582

Qy          53 ArgCysGlySer 56
               |||||||||:::
Db        8583 CGCTGCGGTACT 8594

ESULT 14
US-09-221-017B-894/c
 Sequence 894, Application US/09221017B
 Patent No. 6444799
  GENERAL INFORMATION:
    APPLICANT:  Ross, Bruce C.
    TITLE OF INVENTION:  P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
    NUMBER OF SEQUENCES:  1120
    CORRESPONDENCE ADDRESS:
      ADDRESSEE:  MORRISON & FOERSTER
      STREET:  755 PAGE MILL ROAD
      CITY:  Palo Alto
      STATE:  CA
      COUNTRY:  USA
      ZIP:  94304-1018
    COMPUTER READABLE FORM:
      MEDIUM TYPE:  Diskette
      COMPUTER:  IBM Compatible
      OPERATING SYSTEM:  Windows
      SOFTWARE:  FastSEQ for Windows Version 2.0b
    CURRENT APPLICATION DATA:
      APPLICATION NUMBER:  US/09/221,017B
      FILING DATE:  23-DEC-1998
      CLASSIFICATION:
    PRIOR APPLICATION DATA:
      APPLICATION NUMBER:  PP1182
      FILING DATE:  31-DEC-1997
    PRIOR APPLICATION DATA:
      APPLICATION NUMBER:  PP1546
      FILING DATE:  30-JAN-1998
    PRIOR APPLICATION DATA:
      APPLICATION NUMBER:  PP2911
      FILING DATE:  09-APR-1998
    PRIOR APPLICATION DATA:
      APPLICATION NUMBER:  PCT/AU98/01023
      FILING DATE:  10-DEC-1998
    ATTORNEY/AGENT INFORMATION:
      NAME:  Monroy, Gladys H
;      REGISTRATION NUMBER:  42,430
;      REFERENCE/DOCKET NUMBER:  27943.20021.00
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650 813 5600
;      TELEFAX:  650 494-0792
;      TELEX:  706141
;  INFORMATION FOR SEQ ID NO:  894:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2073 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  circular
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  UNKNOWN
;    ORIGINAL SOURCE:
;      ORGANISM:  PORPHYROMONAS GINGIVALIS
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..2073
US-09-221-017B-894

Alignment Scores:
Pred. No.:               5.e-06        Length:          2073
Score:                   69.0          Matches:         24
Percent Similarity:      37.50%        Conservative:    4
Best Local Similarity:   32.00%        Mismatches:      12
Query Match:             21.44%        Indels:          35
DB:                      4             Gaps:            6

US-09-980-043-4 (1-56) x US-09-221-017B-894 (1-2073)

Qy           7 ProIleCysHisTh---------  --          -------CysGlyGluGluValGlyHis 38
               |||   |||:|||                        :|||||:|:::::|||||:
Db         483 CCTCGGTGCCATAC-GGGCAGCTCCAAGGTAAGTACTTATGTGGAGAGCGGATCGGGCAT 424

Qy          19 AspAlaAsn-----  ---------  --          -----GlyGluLeuPhe--------- 25
                             :                        |:|:::|:|||
Db         423 TGCCGAAGGAGGAT-AACGGGTTTGACGGCATTGGTTACGGGCAGCTTTTTCCTGTTGGC 364

Qy          26 -- ------ValAla------------CysHisGluCysSerTyrProMetCysLysSer 58
                  |:|||              :||      ||||:||||        :||   ||'
Db         363 TCTTTTTCTGTCGC-CCTGTTTCTCCTTGTC-CAGGTGCAGCTACGACAGGTGC---TCT 307

Qy          59 CysPheGluPheGlnIleLysGluGlyArgLysValCysLeuArg 53
               |||                  |||' |   :::      |||
Db         306 TGT ---------  ------AATGGTAGGGGTATTTATGCTCGG 280

RESULT 15
US-09-107-532A-1732/c
; Sequence 1732, Application US/09107532A
; Patent No. 6583275
;   GENERAL INFORMATION:
;        APPLICANT: Lynn A Doucette-Stamm  and David Bush
;        TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
;                            ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
;        NUMBER OF SEQUENCES: 7310
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: GENOME THERAPEUTICS CORPORATION
;             STREET: 100 Beaver Street
;             CITY: Waltham
;             STATE: Massachusetts
;             COUNTRY: USA
;             ZIP: 02354
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: CD/ROM ISO9660
;             COMPUTER: PC
;             OPERATING SYSTEM: <Unknown>
;             SOFTWARE: ASCII
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/107,532A
;             FILING DATE: 30-Jun-1998
;        PRIOR APPLICATION DATA:
```

```
;              APPLICATION NUMBER: 60/085,598
;              FILING DATE: 14 May 1998
;              APPLICATION NUMBER: 60/051571
;              FILING DATE: July 2, 1997
;         ATTORNEY/AGENT INFORMATION:
;              NAME: Ariniello, Pamela Deneke
;              REGISTRATION NUMBER: 40,489
;              REFERENCE/DOCKET NUMBER: GTC-012
;         TELECOMMUNICATION INFORMATION:
;              TELEPHONE: (781)893-5007
;              TELEFAX: (781)893-8277
;    INFORMATION FOR SEQ ID NO: 1732:
;         SEQUENCE CHARACTERISTICS:
;              LENGTH: 861 base pairs
;              TYPE: nucleic acid
;              STRANDEDNESS: double
;              TOPOLOGY: circular
;         MOLECULE TYPE: DNA (genomic)
;         HYPOTHETICAL: NO
;         ANTI-SENSE: NO
;         ORIGINAL SOURCE:
;              ORGANISM: Enterococcus faecium
;         FEATURE:
;              NAME/KEY:  misc_feature
;              LOCATION: (B) LOCATION 1..861
;         SEQUENCE DESCRIPTION: SEQ ID NO. 1732:
US 09-107-532A-1732

Alignment Scores:
Pred. No.:              4.35          Length:          861
Score:                  68.00         Matches:         16
Percent Similarity:     43.14%        Conservative.    6
Best Local Similarity.  31.37%        Mismatches:      21
Query Match:            20.99%        Indels:          8
DB:                     4             Gaps:            2

US 09-980-043 4 (1-56) x US-09-107-532A-1732 (1 861)

Qy          9 CysHisThrCysGlyGluGlnVal  ---------------   GlyHisAspAlaAsn 21
              :|||||:| ||||   :::   :|.                  ||.|||| |   :||
Db        213 TGTCACACTTGCTCCGATTCCGTAAGAATTACCTGCAACACATGCGGTCATGACAGAAAC 154

Qy         22 GlyGluLeuPheValAlaCysHisGluCysSerTyrProMetCysLysSerCysPheGlu 41
                 :::   ::::::|||    ||||    :::                 :||
Db        153 ---ACCGTAAACATTTCTTGTAAAGAATTTACTTATACTTTTTGCAICCCTGCGCTTCC 97

Qy         42 PheGluIleLysGluGlyArgGlyVa
                |||          :   |:|
Db         96 ATCGCTATCTGATCCTGCGATTGTCACTTTTTA
```

Search completed: October 31, 2003, 17:45:3-
Job time : 13.882 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame plus_p2n model

Run on: October 31, 2003, 09:25:59 ; Search time 41.2147 Seconds
(without alignments)
667.830 Million cell updates/sec

Title: US-09-980-043-4
Perfect score: 324
Sequence: 1 MMESGAPICHTCGEQVGHDA.....KSCHEPPIKEGRKVCLRCGS 56

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09980043/runat_29102003_163852_21691/app_query.fasta_1.1358
-DB=N_Geneseq_19Jun03 -QFMT=fastap SUFFIX=rng MINMATCH=0.1 LOOPCL=0
-LOOPEXT=0 UNITS=bits -START=1 -END=-1 MATRIX=blosum... TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
MODE=LOCAL OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09980043_sCGN_1_1_458_@runat_29102003_163852_21691 -NCPU=6 -ICPU=3
NO_MMAP LARGEQUERY NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 DELOP=6 DELEXT=7

Database : N_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqn-em:/NA1987.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-em:/NA1988.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-em:/NA1989.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-em:/NA1990.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-em:/NA1991.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-em:/NA1992.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-em:/NA1993.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-em:/NA1994.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-em:/NA1995.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-em:/NA1996.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-em:/NA1997.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-em:/NA1998.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-em:/NA1999.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-em:/NA2000.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-em:/NA2001A.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-em:/NA2001B.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-em:/NA2002.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-em:/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	324	100.0	1010	22	AAC65449	Populus tremuloide
2	319	98.5	3232	22	AAC65448	Populus tremuloide
3	264	81.5	3328	19	AAV34432	Cotton cellulose s
4	256	79.0	3207	19	AAV08372	Cellulose synthase
5	233	71.9	664	21	AAA67127	Eucalyptus grandis
6	233	71.9	856	21	AAA67079	Eucalyptus grandis
7	233	71.9	3747	21	AAA67145	Eucalyptus grandis
8	165	50.9	736	21	AAA67115	Pinus radiata cell
9	165	50.9	3851	21	AAA67114	Pinus radiata cell
10	164	50.6	3311	19	AAV08373	Cellulose synthase
11	162	50.0	3753	21	AAZ99515	DNA encoding a mai
12	162	50.0	3753	21	AAZ99530	DNA encoding a mai
13	162	50.0	3780	21	AAZ99497	DNA encoding a mai
14	161	49.7	1741	19	AAV06569	Rice cellulose syn
15	161	49.7	3725	21	AAZ99500	DNA encoding a mai
16	161	49.7	3725	21	AAZ99506	DNA encoding a mai
17	161	49.7	3725	21	AAZ99521	DNA encoding a mai
18	158	48.8	3568	21	AAZ99491	DNA encoding a mai
19	158	48.8	3776	21	AAZ58263	Corn cellulose syn
20	156	48.1	1086	19	AAV08374	Cellulose synthase
21	154	47.5	3517	21	AAZ58268	Soybean cellulose
22	153.5	47.4	3786	21	AAZ58265	Corn cellulose syn
23	153.5	47.4	3813	21	AAZ99509	DNA encoding a mai
24	153.5	47.4	3813	21	AAZ99524	DNA encoding a mai
25	152	46.9	3603	19	AAV06565	Arabidopsis cellul
26	152	46.9	3673	19	AAV06568	Arabidopsis cellul
27	152	46.9	7234	22	AAC83247	Cellulose synthase
28	152	46.9	8411	19	AAV06563	Arabidopsis cellul
29	150	46.3	3936	21	AAZ58266	Corn cellulose syn
30	150	46.3	3969	21	AAZ99503	DNA encoding a mai
31	150	46.3	3969	21	AAZ99518	DNA encoding a mai
32	148	45.7	3255	24	ABZ12754	Arabidopsis thalia
33	148	45.7	3746	21	AAZ99512	DNA encoding a mai
34	148	45.7	3746	21	AAZ99527	DNA encoding a mai
35	148	45.7	3773	21	AAZ99494	DNA encoding a mai
36	148	45.7	3799	24	AAS16458	Corn cDNA encoding
37	148	45.7	3828	19	AAV06566	Arabidopsis cellul
38	148	45.7	5009	19	AAV06564	Arabidopsis cellul
39	146	45.1	3704	21	AAZ99533	DNA encoding a mai
40	140	43.2	641	24	ABQ66034	Arabidopsis thalia
41	139	42.9	3255	24	ABZ14725	Arabidopsis thalia
42	139	42.9	3444	22	AAC83798	Arabidopsis thalia
43	138	42.6	3198	21	AAC49550	Arabidopsis thalia
44	138	42.6	3614	19	AAV06567	Arabidopsis cellul
45	133	41.0	264	21	AAA31474	Plant microsatelli

ALIGNMENTS

RESULT 1
AAC65449
ID AAC65449 standard; DNA; 1010 BP.
XX
AC AAC65449;
XX
DT 07-MAR-2001 (first entry)
XX
DE Populus tremuloides CelA 5' flanking region.
XX
KW Aspen; CelA; cellulose synthase; arabidopsis; wood;
KW mechanical stress response element; MSRE; cellulose biosynthesis; ds.
XX
OS Populus tremuloides.
XX
PN WO200071670-A2.
XX

```
D   30-NOV-2000.
X
F   18-MAY-2000; 2000WO-US13637.
X
R   21-MAY-1999;   99US-0135280.
X
A   (UNMT ) UNIV MICHIGAN TECHNOLOGICAL.
X
I   Chiang VL,  Wu L,  Chandrashekhar JP,  Carraway DT;
X
R   WPI; 2001-032025/04.
R   P-PSDB; AAB37894.
X
T   Novel polynucleotide encoding cellulose synthase useful for enhancing
T   cellulose biosynthesis, modifying lignin biosynthesis, and for
T   accelerating growth of plants and for upregulating cellulose synthase
T   levels -
X
S   Claim 2; Fig 4; 65pp; English.
X
C   The present sequence is given in a specification relating to
C   polynucleotides encoding cellulose synthase enzymes from Populus
C   tremuloides and Arabidopsis. The cellulose synthases are designated
C   PtCelA and AraxCelA respectively. The polynucleotides may be introduced
C   into plants where they are useful for altering the growth of a plant.
C   Vectors comprising a cellulose synthase promoter, or a fragment of the
C   promoter containing one or more mechanical stress response elements
C   (MSRE), operably linked to a cellulose synthase gene are useful for
C   causing stress-induced gene expression in a plant cell. A cDNA encoding a
C   protein that binds to a positive MSRE may be delivered into the plant to
C   increase expression of cellulose synthase. A cDNA in an antisense
C   orientation may be delivered to reduce expression of cellulose synthase.
C   The polynucleotides enable the development of transgenic tree varieties
C   having increased cellulose content, decreased lignin content and
C   therefore improved wood fibre characteristics. The cellulose synthase
C   promoters are useful for synthesising greater quantities of
C   high-crystalline cellulose in plant and preferably in trees, which
C   permits an increase in proportion of cellulose in transgenic plants,
C   greater strength of juvenile wood and fiber and acceleration of overall
C   growth rate.
XX
Q   Sequence 1010 BP; 289 A; 212 C; 216 G; 293 T; 0 other;

Alignment Scores:
Pred. No.:                2.76e-32        Length:         1010
Score:                    324.00          Matches:        56
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       22              Gaps:           0

US-09-980-043-4 (1-56) x AAC65449 (1-1010)

Qy          1 MetMetGluSerGlyAlaProIleCysHisThrCysGlyGluGlnValGlyHisAspAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 ATGATGGAATCTGGGGCTCCTATATGCCATACCTGTGGTGAACAGGTGGGGCATGATGCA 900

Qy         21 AsnGlyGluLeuPheValAlaCysHisGluCysSerTyrProMetCysLysSerCysPhe 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AATGGGGAGCTATTTGTGGCTTGCCATGAGTGTAGCTATCCCATGTGCAAGTCTTGTTTC 960

Qy         41 GluPheGluIleLysGluGlyArgLysValCysLeuArgCysGlySer 56
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GAGTTTGAAATCAAAGAGGGCCGGAAAGTTTGCTTGCGGTGTGGCTCG 1008

RESULT 2
AAC65448
ID   AAC65448 standard; cDNA; 3232 BP.
XX
AC   AAC65448;
XX
DT   07-MAR-2001  (first entry)
XX
DE   Populus tremuloides cell cDNA
XX
KW   Aspen; CelA; cellulose synthase; Arabidopsis; wood;
KW   mechanical stress response element; MSRE; cellulose biosynthesis; ss.
XX
OS   Populus tremuloides
XX
PN   WO200071670-A2.
XX
PD   30-NOV-2000.
XX
PF   18-MAY-2000; 2000WO-US13637.
XX
PR   21-MAY-1999;   99US-0135280.
XX
PA   (UNMT ) UNIV MICHIGAN TECHNOLOGICAL.
XX
PI   Chiang VL,  Wu L,  Chandrashekhar JP,  Carraway DT;
XX
DR   WPI; 2001-032025/04.
DR   P-PSDB; AAB37894.
XX
PT   Novel polynucleotide encoding cellulose synthase useful for enhancing
PT   cellulose biosynthesis, modifying lignin biosynthesis, and for
PT   accelerating growth of plants and for upregulating cellulose synthase
PT   levels -
XX
PS   Claim 2; Fig 4; 65pp; English.
XX
CC   The present sequence is given in a specification relating to
CC   polynucleotides encoding cellulose synthase enzymes from Populus
CC   tremuloides and Arabidopsis. The cellulose synthases are designated
CC   PtCelA and AraxCelA respectively. The polynucleotides may be introduced
CC   into plants where they are useful for altering the growth of a plant.
CC   Vectors comprising a cellulose synthase promoter, or a fragment of the
CC   promoter containing one or more mechanical stress response elements
CC   (MSRE), operably linked to a cellulose synthase gene are useful for
CC   causing stress-induced gene expression in a plant cell. A cDNA encoding a
CC   protein that binds to a positive MSRE may be delivered into the plant to
CC   increase expression of cellulose synthase. A cDNA in an antisense
CC   orientation may be delivered to reduce expression of cellulose synthase
CC   The polynucleotides enable the development of transgenic tree varieties
CC   having increased cellulose content, decreased lignin content and
CC   therefore improved wood fibre characteristics. The cellulose synthase
CC   promoters are useful for synthesising greater quantities of
CC   high-crystalline cellulose in plant and preferably in trees, which
CC   permits an increase in proportion of cellulose in transgenic plants,
CC   greater strength of juvenile wood and fiber and acceleration of overall
CC   growth rate.
XX
SQ   Sequence 3232 BP; 895 A; 668 C; 772 G; 897 T; 0 other;

Alignment Scores:
Pred. No.:                5.44e-31        Length:         3232
Score:                    319.00          Matches:        55
Percent Similarity:       98.21%          Conservative:   0
Best Local Similarity:    98.21%          Mismatches:     1
Query Match:              98.46%          Indels:         0
DB:                       22              Gaps:           0

US-09-980-043-4 (1-56) x AAC65448 (1-3232)

Qy          1 MetMetGluSerGlyAlaProIleCysHisThrCysGlyGluGlnValGlyHisAspAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         69 ATGATGGAATCTGGGGCTCCTATATGCCATACCTGTGGTGAACAGGTGGGGCATGATGCA 128

Qy         21 AsnGlyGluLeuPheValAlaCysHisGluCysSerTyrProMetCysLysSerCysPhe 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        129 AATGGGGAGCTATTTGTGGCTTGCCATGAGTGTAGCTATCCCATGTGCAAGTCTTGTTTC 188

Qy         41 GluPheGluIleLysGluGlyArgLysValCysLeuArgCysGlySer 56
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        189 GAGTTTGAAATCAATGAGGGCCGGAAAGTTTGCTTGCGGTGTGGCTCG 236
```

```
RESULT 3
AAV34432
ID   AAV34432 standard; cDNA; 3328 BP.
XX
AC   AAV34432;
XX
DT   28-SEP-1998  (first entry)
XX
DE   Cotton cellulose synthase CelA1 cDNA.
XX
KW   Cellulose synthase; CelA1 gene; cotton; forest tree; plant fibre;
KW   wood; transgenic plant; ds.
XX
OS   Gossypium hirsutum Acala SJ-2.
XX
FH   Key             Location/Qualifiers
FT   CDS             179..3103
FT                   /*tag=  a
XX
PN   WO9818949-A2.
XX
PD   07-MAY-1998.
XX
PF   29-OCT-1997;   97WO-US19529.
XX
PR   29-OCT-1996;   96US-0029987.
XX
PA   (CALJ ) CALGENE INC.
XX
PI   Delmer DP,  Pear JR,  Stalker DM;
XX
DR   WPI; 1998-272238/24.
DR   P-PSDB; AAW60039.
XX
PT   New isolated plant cellulose synthesis enzyme DNA  is used to
PT   develop products for producing modified plant fibre products,
PT   particularly for modifying cotton fibre and wood quality
XX
PS   Claim 9; Fig 6A-E; 49pp; English.
XX
CC   This full-length cDNA sequence corresponds to the cotton CelA1 gene
CC   transcript and encodes a 973-amino acid cellulose synthase (see
CC   AAW60039).  2 cDNA clones that exhibited small blocks of amino acid
CC   homology to proteins encoded by bacterial CelA genes were isolated
CC   from an unamplified cDNA library prepared from mRNA of fibres of
CC   cotton Acala SJ-2 harvested at 24 dpa (time of maximal secondary
CC   wall cellulose synthesis).  These 2 clones were partially homologous
CC   at the nucleotide and amino acid levels and were designated CelA1
CC   and CelA2 (see AAV34433), respectively.  The 2 cotton CelA genes are
CC   both expressed in developing fibre at approximately 500 times their
CC   level of expression in other cotton tissues, and constitute
CC   approximately 1-2% of the 24 dpa fibre mRNA.  The CelA1 gene
CC   promoter region (see AAV34434) is also claimed.  The isolated cDNA
CC   clones are used in claimed methods for modifying fibre phenotype
CC   in a cotton plant and for modifying wood quality phenotype in a
CC   forest tree species.  Lignin can be reduced by over-expressing the
CC   cellulose synthase product and diverting production into cellulose.
XX
SQ   Sequence 3328 BP; 914 A; 690 C; 763 G; 961 T; 0 other;

Alignment Scores:
Pred. No.:                7.14e-24        Length:         3328
Score:                    264.00          Matches:        42
Percent Similarity:       89.29%          Conservative:   3
Best Local Similarity:    75.00%          Mismatches:     6
Query Match:              81.48%          Indels:         5
DB:                       19              Gaps:           5

US-09-980-043-4 (1-56) x AAV34432 (1-3328)

Qy          1 MetMetGluSerGlyAlaProIleCysHisThrCysGlyGluGlnValGlyHisAspAla 20
              | ||||:| |||| '   |. :::|  :|||. ||||| |:|   ||'|:|   :::

Db        179 ATGATGGAATCTGGGGTTCCTGTTTGCCACACTTGTGGTGAACATGTTGGGTTGAATGTT 238

Qy         21 AsnGlyGluLeuPheValAlaCysHisGluCysSerTyrProMetCysLysSerCysPhe 40
              |||||||||   ||||||||||||||||||||:::::|||:::|||||||||||||||||

Db        239 AATGGTGAACCTTTTGTGGCTTGCCATGAATGTAATTTCCCTATTTGTAAGAGTTGTTTT 298

Qy         41 GluPheGluIleLysGluGlyArgLysValCysLeuArgCysGlySer 56
              |||::::::::::|||||||||||||||   ||||||||||||||||||

Db        299 GAGTATGATCTTAAGGAAGGACGAAAAGCTTGCTTGCGTTGTGGTAGT 346

RESULT 4
AAV08372
ID   AAV08372 standard; cDNA to mRNA; 3207 BP.
XX
AC   AAV08372;
XX
DT   08-FEB-1999  (first entry)
XX
DE   Cellulose synthase PcsA1 coding sequence.
XX
KW   Cellulose synthase; cotton; ds.
XX
OS   Gossypium hirsutum.
XX
FH   Key             Location/Qualifiers
FT   CDS             77..3001
FT                   /*tag=  a
XX
PN   EP875575-A2.
XX
PD   04-NOV-1998.
XX
PF   31-MAR-1998;   98EP-0302489.
XX
PR   01-APR-1997;   97JP-0083133.
XX
PA   (NISN ) NISSHINBO IND INC.
XX
PI   Aotsuka S,  Hasegawa O,  Hayashi T,  Ihara Y;
XX
DR   WPI; 1998-559440/48.
DR   P-PSDB; AAW73308.
XX
PT   New DNA encoding cotton cellulose synthase proteins - useful for
PT   controlling cellulose synthesis in prokaryotic or eukaryotic cells
XX
PS   Example 5; Page 14-18; 51pp; English.
XX
CC   This sequence encodes a cellulose synthase of the invention. The DNA can
CC   be used for controlling cellulose synthesis in a cell by introducing the
CC   DNA into the cell, expressing RNA with a nucleotide sequence homologous
CC   to the DNA or a nucleotide sequence complementary to the DNA. The cells
CC   are prokaryotic cells i.e. acetobacterium and/or eukaryotic cells such as
CC   yeasts i.e. Saccharomyces spp., plant cells (cotton plants) and cultured
CC   cells of mammals etc.
XX
SQ   Sequence 3207 BP; 883 A; 662 C; 723 G; 939 T; 0 other;

Alignment Scores:
Pred. No.:                7.35e-23        Length:         3207
Score:                    256.00          Matches:        40
Percent Similarity:       89.09%          Conservative:   9
Best Local Similarity:    72.73%          Mismatches:     6
Query Match:              79.01%          Indels:         0
DB:                       19              Gaps:           0

US-09-980-043-4 (1-56) x AAV08372 (1-3207)

Qy          1 MetMetGluSerGlyAlaProIleCysHisThrCysGlyGluGlnValGlyHisAspAla 20
              ||:|||:|||||||   |||:::|||||||||||||||||||   ||||||   :::

Db         77 ATGATGGAATCTGGGGTTCCTGTTTGCCACACTTGTGGTGAACATGTTGGGTTGAATGTT 136

Qy         21 AsnGlyGluLeuPheValAlaCysHisGluCysSerTyrProMetCysLysSerCysPhe 40
```

```
Db        137 AATGGTGAACCCTTTGTGGCTTGCCATGAATGTAATTTCCCTATTTGTAAGAGTTGTTTT 196

Qy         41 GluPheGluIleLysGluGlyArgLysValCysLeuArgCysGly 55

Db        197 GAGTATGATCTTAAGGAAGGACAAAAAGCTTGCTTGCGTTGTGGT 241

ESULT 5
AA67127
D   AAA67127 standard; DNA; 664 BP.
X
C   AAA67127;
X
T   31-OCT-2000  (first entry)
X
E   Eucalyptus grandis cellulose synthase nucleotide sequence SEQ ID NO:107.
X
W   Eucalyptus grandis; pinus radiata; Monterey pine; modification;
W   plant cell wall; polysaccharide; polysaccharide biosynthetic pathway;
W   transgenic plant; ds.
X
S   Eucalyptus grandis.
X
N   WO200022092-A2.
X
D   20-APR-2000.
X
F   08-OCT-1999;   99WO-NZ00169.
X
R   13-OCT-1998;   98US-0170862.
R   11-AUG-1999;   99US-0148426.
X
A   (GENE-) GENESIS RES & DEV CORP LTD.
A   (FLET-) FLETCHER CHALLENGE FORESTS LTD.
X
I   Bloksberg LN;
X
R   WPI; 2000-339328/29.
R   P-PSDB; AAB16320.
X
T   New genes encoding proteins involved in a plant polysaccharide
T   biosynthetic pathway, useful for modulating or altering the
T   polysaccharide content, composition or structure of the plant
X
S   Claim 1; Page 92; 301pp; English.
X
C   The present invention describes isolated polynucleotides (PN) comprising
C   a sequence selected from one of 835 nucleotide sequences given in
C   AAA67073 to AAA67907, their (reverse) complements, sequences producing
C   an Expectation (E) value of 0.01 or less compared to the 835 sequences,
C   sequences at least 50% identical to them, 200, 100, 40 or 20-mers of the
C   835 sequences or sequences that are degenerately equivalent or allelic
C   to the 835 sequences. The polynucleotides are used to modify the
C   activity of a polypeptide involved in a polysaccharide biosynthetic
C   pathway in the plant. They are especially used to modulate or alter the
C   polysaccharide content, composition or structure of the plant. AAB16268
C   to AAB16340 are proteins encoded by some of the polynucleotide sequence
C   given in the present invention.
X
Q   Sequence 664 BP; 201 A; 136 C; 182 G; 145 T; 0 other;

Alignment Scores:
Pred. No.:              9.08e-21        Length:         664
Score:                  233.00          Matches:        38
Percent Similarity:     83.93%          Conservative:   9
Best Local Similarity:  67.86%          Mismatches:     9
Query Match:            71.91%          Indels:         0
DB:                     21              Gaps:           0

US-09-980-043-4 (1-56) x AAA67127 (1-664)

Qy          1 MetMetGluSerGlyAlaProIleCysHisThrCysGlyGluGlnValGlyHisAspAla 20

Db        111 ATGATGGAAICCGAITTCCCTTGCAACATTGCGGAGAGGCTGTTGGGGTTGATGAG 170

Qy         21 AsnGlyGluLeuPheValAlaCysHisGluCysSerTyrProMetCysLysSerCysPhe 40

Db        171 AAAGGCGAGGTCTTCGTGGCTTGTCAAGAGTGCAACTTCGCCATTTGCAAGGCTTGTGTC 230

Qy         41 GluPheGluIleLysGluGlyArgLysValCysLeuArgCysGlySer 56

Db        231 GAATATGAGATTAAGGAAGGAAGAAAAAGCGTGTTGCGCTGTGGCACT 278

RESULT 6
AAA67079
ID   AAA67079 standard; DNA; 856 BP
XX
AC   AAA67079;
XX
DT   31-OCT-2000  (first entry)
XX
DE   Eucalyptus grandis cellulose synthase nucleotide sequence SEQ ID NO:7.
XX
KW   Eucalyptus grandis; pinus radiata; Monterey pine; modification;
KW   plant cell wall; polysaccharide; polysaccharide biosynthetic pathway;
KW   transgenic plant; ds.
XX
OS   Eucalyptus grandis
XX
PN   WO200022092-A2.
XX
PD   20-APR-2000.
XX
PF   08-OCT-1999;   99WO-NZ00169.
XX
PR   13-OCT-1998;   98US-0170862.
PR   11-AUG-1999;   99US-0148426.
XX
PA   (GENE-) GENESIS RES & DEV CORP LTD
PA   (FLET-) FLETCHER CHALLENGE FORESTS LTD
XX
PI   Bloksberg LN;
XX
DR   WPI; 2000-339328/29.
DR   P-PSDB; AAB16274.
XX
PT   New genes encoding proteins involved in a plant polysaccharide
PT   biosynthetic pathway, useful for modulating or altering the
PT   polysaccharide content, composition or structure of the plant
XX
PS   Claim 1; Page 40; 301pp; English.
XX
CC   The present invention describes isolated polynucleotides (PN) comprising
CC   a sequence selected from one of 835 nucleotide sequences given in
CC   AAA67073 to AAA67907, their (reverse) complements, sequences producing
CC   an Expectation (E) value of 0.01 or less compared to the 835 sequences,
CC   sequences at least 50% identical to them, 200, 100, 40 or 20-mers of the
CC   835 sequences or sequences that are degenerately equivalent or allelic
CC   to the 835 sequences. The polynucleotides are used to modify the
CC   activity of a polypeptide involved in a polysaccharide biosynthetic
CC   pathway in the plant. They are especially used to modulate or alter the
CC   polysaccharide content, composition or structure of the plant. AAB16268
CC   to AAB16340 are proteins encoded by some of the polynucleotide sequence
CC   given in the present invention.
XX
SQ   Sequence 856 BP; 247 A; 179 C; 226 G; 204 T; 0 other;

Alignment Scores:
Pred. No.:              1.26e-20        Length:         856
Score:                  233.00          Matches:        38
Percent Similarity:     83.93%          Conservative:   9
Best Local Similarity:  67.86%          Mismatches:     9
Query Match:            71.91%          Indels:         0
DB:                     21              Gaps:           0

US-09-980-043-4 (1-56) x AAA67079 (1-856)
```

```
Qy          1 MetMetGluSerGlyAlaProIleCysHisThrCysGlyGluGlnValGlyHisAspAla 20
              |||||||||||||||   |||:::|||::: |||:|||||: |    ||||:   :||
Db        103 ATGATGGAATCCGGGGTTCCCCTGTGCAACACTTGCGGAGAGGCTGTTGGGGTTGATGAG 162

Qy         21 AsnGlyGluLeuPheValAlaCysHisGluCysSerTyrProMetCysLysSerCysPhe 40
              |||:|||:::|||||:|||||    ||| ||:::::   :::|:||||:::|||
Db        163 AAAGGCGAGGTCTTCGTGGCTTGTCAAGAGTGCAACTTCGCCATTTGCAAGGCTTGTGTC 222

Qy         41 GluPheGluIleLysGluGlyArgLysValCysLeuArgCysGlySer 56
              ||:::||:|||||:|||||||||||||:    :|||||| ||| ||:::
Db        223 GAATATGAGATTAAGGAAGGAAGAAAAGCGTGCTTGCGCTGTGGCACT 270

RESULT 7
AAA67145
ID   AAA67145 standard; DNA; 3747 BP.
XX
AC   AAA67145;
XX
DT   31-OCT-2000  (first entry)
XX
DE   Eucalyptus grandis cellulose synthase nucleotide sequence SEQ ID NO:141.
XX
KW   Eucalyptus grandis; pinus radiata; Monterey pine; modification;
KW   plant cell wall; polysaccharide; polysaccharide biosynthetic pathway;
KW   transgenic plant; ds.
XX
OS   Eucalyptus grandis.
XX
PN   WO200022092-A2.
XX
PD   20 APR 2000.
XX
PF   08-OCT-1999;   99WO-NZ00169.
XX
PR   13 OCT 1998;   98US-0170862.
PR   11-AUG-1999;   99US-0148426.
XX
PA   (GENE-) GENESIS RES & DEV CORP LTD.
PA   (FLET-) FLETCHER CHALLENGE FORESTS LTD
XX
PI   Bleksberg LN;
XX
DR   WPI; 2000-339328/29
DR   P-PSDB; AAB16338.
XX
PT   New genes encoding proteins involved in a plant polysaccharide
PT   biosynthetic pathway, useful for modulating or altering the
PT   polysaccharide content, composition or structure of the plant -
XX
PS   Claim 1; Page 106-107; 301pp; English.
XX
CC   The present invention describes isolated polynucleotides (PN) comprising
CC   a sequence selected from one of 835 nucleotide sequences given in
CC   AAA67073 to AAA67907, their (reverse) complements, sequences producing
CC   an Expectation (E) value of 0.01 or less compared to the 835 sequences,
CC   sequences at least 50% identical to them, 200, 100, 40 or 20-mers of the
CC   835 sequences or sequences that are degenerately equivalent or allelic
CC   to the 835 sequences. The polynucleotides are used to modify the
CC   activity of a polypeptide involved in a polysaccharide biosynthetic
CC   pathway in the plant. They are especially used to modulate or alter the
CC   polysaccharide content, composition or structure of the plant. AAB16268
CC   to AAB16340 are proteins encoded by some of the polynucleotide sequence
CC   given in the present invention.
XX
SQ   Sequence 3747 BP; 976 A; 822 C; 935 G; 1014 T; 0 other;

Alignment Scores:
Pred. No.:              8.37e-20        Length:         3747
Score:                  233.00          Matches:        39
Percent Similarity:     83.93%          Conservative:   0
Best Local Similarity:  67.86%          Mismatches:     9
Query Match:            71.91%          Indels:         5
DB:                     21              Gaps:           0

US-09-980-043-4 (1-56) x AAA67145 (1-3747)

Qy          1 MetMetGluSerGlyAlaProIleCysHisThrCysGlyGluGlnValGlyHisAspAla 20
              |||||||||||||||   | |:::|||:::|||||||||||   ||||||   |||
Db        511 ATGATGGAATCCGGGGTTCCCCTGTGCAACACTTGCGGAGAGGCTGTTGGGGTTGATGAG 570

Qy         21 AsnGlyGluLeuPheValAlaCysHisGluCysSerTyrProMetCysLysSerCysPhe 40
              ||||||:::|||||||:||||   ||||||::::::   :::|||||:::|||
Db        571 AAAGGCGAGGTCTTCGTGGCTTGTCAAGAGTGCAACTTCGCCATTTGCAAGGCTTGTGTC 630

Qy         41 GluPheGluIleLysGluGlyArgLysValCysLeuArgCysGlySer 56
              |||:::|||||||||||||:|||||||   |||||||||||||||:::
Db        631 GAATATGAGATTAAGGAAGGAAGAAAAGCGTGCTTGCGCTGTGGCACT 678

RESULT 8
AAA67115
ID   AAA67115 standard; DNA; 736 BP.
XX
AC   AAA67115;
XX
DT   31-OCT-2000  (first entry)
XX
DE   Pinus radiata cellulose synthase nucleotide sequence SEQ ID NO:70.
XX
KW   Eucalyptus grandis; pinus radiata; Monterey pine; modification;
KW   plant cell wall; polysaccharide; polysaccharide biosynthetic pathway;
KW   transgenic plant; ds.
XX
OS   Pinus radiata.
XX
PN   WO200022092-A2.
XX
PD   20-APR-2000.
XX
PF   08-OCT-1999;   99WO-NZ00169.
XX
PR   13-OCT-1998;   98US-0170862.
PR   11-AUG-1999;   99US-0148426.
XX
PA   (GENE ) GENESIS RES & DEV CORP LTD.
PA   (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI   Bleksberg LN.
XX
DR   WPI; 2000-339328/29.
DR   P-PSDB; AAB16308.
XX
PT   New genes encoding proteins involved in a plant polysaccharide
PT   biosynthetic pathway, useful for modulating or altering the
PT   polysaccharide content, composition or structure of the plant -
XX
PS   Claim 1; Page 72; 301pp; English.
XX
CC   The present invention describes isolated polynucleotides (PN) comprising
CC   a sequence selected from one of 835 nucleotide sequences given in
CC   AAA67073 to AAA67907, their (reverse) complements, sequences producing
CC   an Expectation (E) value of 0.01 or less compared to the 835 sequences,
CC   sequences at least 50% identical to them, 200, 100, 40 or 20-mers of the
CC   835 sequences or sequences that are degenerately equivalent or allelic
CC   to the 835 sequences. The polynucleotides are used to modify the
CC   activity of a polypeptide involved in a polysaccharide biosynthetic
CC   pathway in the plant. They are especially used to modulate or alter the
CC   polysaccharide content, composition or structure of the plant. AAB16268
CC   to AAB16340 are proteins encoded by some of the polynucleotide sequence
CC   given in the present invention.
XX
SQ   Sequence 736 BP; 216 A; 163 C; 210 G; 147 T; 0 other;

Alignment Scores:
Pred. No.:              6.25e-12        Length:         736
Score:                  165.00          Matches:        23
```

```
ercent Similarity:     78.26%          Conservative:   13
est Local Similarity:  50.00%          Mismatches:     10
uery Match:            50.93%          Indels:         0
B:                     21              Gaps:           0

S-09-980-043-4 (1-56) x AAA67115 (1-736)

y           9 CysHisThrCysGlyGluGlnValGlyHisAspAlaAsnGlyGluLeuPheValAlaCys 28
              |||      ||||||::::::::::|||   :::   ||||||||||||||||||||||||
b         157 TGCCAAGTCTGCGGAGATAAAATCGGCTGCAACCCGAATGGAGAACTATTCGTGGCCTGC 216

y          29 HisGluCysSerTyrProMetCysLysSerCysPheGluPheGluIleLysGluGlyArg 48
              :::|||||   :::|||:::|||:::   |||:::|||:::|||   |||:::|||
b         217 AACGAGTGTGGATTCCCTGTGTGTCGTCCCTGTTATGAGTACGAGAGAAAGGATGGGAAC 276

y          49 LysValCysLeuArgCys 54
              :::   |||   :::|||
b         277 CGGTGCTGCCCTCAGTGC 294

ESULT 9
AA67114
D   AAA67114 standard; DNA; 3851 BP.
X
C   AAA67114;
X
T   31-OCT-2000  (first entry)
X
E   Pinus radiata cellulose synthase nucleotide sequence SEQ ID NO:69.
X
W   Eucalyptus grandis; pinus radiata; Monterey pine; modification;
W   plant cell wall; polysaccharide; polysaccharide biosynthetic pathway;
W   transgenic plant; ds.
X
S   Pinus radiata.
X
N   WO200022092-A2.
X
D   20-APR-2000.
X
F   08-OCT-1999;   99WO-NZ00169.
X
R   13-OCT-1998;   98US-0170862.
R   11-AUG-1999;   99US-0148426.
X
A   (GENE-) GENESIS RES & DEV CORP LTD.
A   (FLET-) FLETCHER CHALLENGE FORESTS LTD.
X
I   Bloksberg LN;
X
R   WPI; 2000-339328/29.
R   P-PSDB; AAB16307.
X
T   New genes encoding proteins involved in a plant polysaccharide
T   biosynthetic pathway, useful for modulating or altering the
T   polysaccharide content, composition or structure of the plant
X
S   Claim 1; Page 71-72; 301pp; English.
X
C   The present invention describes isolated polynucleotides (PN) comprising
C   a sequence selected from one of 835 nucleotide sequences given in
C   AAA67073 to AAA67907, their (reverse) complements, sequences producing
C   an Expectation (E) value of 0.01 or less compared to the 835 sequences,
C   sequences at least 50% identical to them, 200, 100, 40 or 20-mers of the
C   835 sequences or sequences that are degenerately equivalent or allelic
C   to the 835 sequences. The polynucleotides are used to modify the
C   activity of a polypeptide involved in a polysaccharide biosynthetic
C   pathway in the plant. They are especially used to modulate or alter the
C   polysaccharide content, composition or structure of the plant. AAB16268
C   to AAB16340 are proteins encoded by some of the polynucleotide sequence
C   given in the present invention.
X
Q   Sequence 3851 BP; 1040 A; 796 C; 944 G; 1071 T; 0 other;

Alignment Scores:
Pred. No.:              5.8e-11         Length:         3851
Score:                  164.00          Matches:        24
Percent Similarity:     78.26%          Conservative:   13
Best Local Similarity:  48.00%          Mismatches:     12
Query Match:            50.93%          Indels:         0
DB:                     21              Gaps:           0

US-09-980-043-4 (1-56) x AAA67114 (1-3851)

Qy          8 IleCysHisThrCysGlyGluGlnValGlyHisAspAlaAsnGlyGluLeuPheValAla 27
              ::: ||      |  |||:     |||  :::   :::|||||||| |||| ||
Db        258 GTCTGCCACATTT  TGTAGGA CTCTTTTCTAACACAGACGGCGAGCTGTTCGTTGCC 317

Qy         28 CysHisGluCysSerTyrProMetCysLysSerCysPheGluPheGluIleLysGluGly 47
              ||:::||| |   :::|  :::   :::  |||:::|||:::|||   :::||  ||
Db        318 TGTAATGAGTGCAGT TTTCCTGTTTGTTCTGGCCTGTGCTATGAGTACGAGAGCGAGAAGGA 377

Qy         48 ArgLysValCysLeuArgCysArgCysGlySer 78
              :::   |  ||  |   |  |  |  :
Db        378 AATGTGTGTCTCAG  AGTGCAGATG  AGC 410

RESULT 10
AAV08373
ID   AAV08373 standard; cDNA; 3311 BP.
XX
AC   AAV08373;
XX
DT   08-FEB-1999  (first entry)
XX
DE   Cellulose synthase PesA2 coding sequence.
XX
KW   Cellulose synthase; cotton; ds.
XX
OS   Gossypium hirsutum
XX
FH   Key             Location/Qualifiers
FT   CDS             22..3144
FT                   /*tag= a
XX
PN   EP875575-A2
XX
PD   04-NOV-1998.
XX
PF   31-MAR-1998;   98EP-0302489
XX
PR   01-APR-1997;   97JP-0083133.
XX
PA   (NISN ) NISSHINBO IND INC.
XX
PI   Aotsuka S, Hasegawa O, Hayashi T, Ihara Y;
XX
DR   WPI; 1998-559440/48.
DR   P-PSDB; AAW73309.
XX
PT   New DNA encoding cotton cellulose synthase proteins - useful for
PT   controlling cellulose synthesis in prokaryotic or eukaryotic cells
XX
PS   Example 5; Page 22-27; 51pp; English.
XX
CC   This sequence encodes a cellulose synthase of the invention. The DNA can
CC   be used for controlling cellulose synthesis in a cell by introducing the
CC   DNA into the cell, expressing RNA with a nucleotide sequence homologous
CC   to the DNA or a nucleotide sequence complementary to the DNA. The cells
CC   are prokaryotic cells i.e. acetobacterium and/or eukaryotic cells such as
CC   yeasts i.e. Saccharomyces spp., plant cells (cotton plants) and cultured
CC   cells of mammals etc.
XX
SQ   Sequence 3311 BP; 900 A; 678 C; 788 G; 945 T; 0 other;

Alignment Scores:
Pred. No.:              5.8e-11         Length:         3311
Score:                  164.00          Matches:        23
```

```
Percent Similarity:     68.52%          Conservative:   14
Best Local Similarity:  42.59%          Mismatches:     17
Query Match:            50.62%          Indels:         0
DB:                     19              Gaps:           2

US-09-980-043-4 (1-56) x AAV08373 (1-3311)

Qy          9 GluSerGlyAlaProIleCysHisThrCysGlyGluGlnVa.GlyHisAspAlaAsnGly 22
              :::|||         |||||:      |||||  :::::::::*|||         ||||||
Db         95 CAATCATCAACGAAAATATGCAGGGTGTGTGGGGATAAGATTGGGCAAAAGGAAAACGGA 154

Qy         23 GluLeuPheValAlaCysHisGluCysSerTyrProMetCysLysSerCysPheGluPhe 42
              :::   |:|||||:||||||   || ::::::|||:::||.:::   :||:::|||:::
Db        155 CAACCGTTCGTGGCTTGTCATGTCTGTGCTTTCCCGGTTTGCCGTCCTTGTTATGAATAT 214

Qy         43 GluIleLysGluGlyArgLysValCysLeuArgCysGlySer 56
              :||      |||||    :::   ||    :::|||   :::
Db        215 GAAAGGAGTGAAGGAAACCAGTGCTGTCCTCAGTGCAATACT 256

RESULT 11
AAZ99515
ID   AAZ99515 standard; DNA; 3753 BP.
XX
AC   AAZ99515;
XX
DT   03-JUL-2000  (first entry)
XX
DE   DNA encoding a maize cellulose synthase.
XX
KW   Maize; cellulose synthase; stalk quality; stand; silage; cellulose;
KW   transgenic plant; plant breeding marker; ss.
XX
OS   Zea mays.
XX
FH   Key             Location/Qualifiers
FT   CDS             184..3405
FT                   /*tag= a
FT                   /product= "cellulose synthase"
FT                   /transl_except= (pos: 916..918, aa: Xaa)
FT                   /note= "no termination codon given; Xaa is an
FT                         unspecified amino acid"
XX
PN   WO200009706-A2.
XX
PD   24-FEB-2000
XX
PF   16-AUG-1999;   99WO-US18760.
XX
PR   17-AUG-1998;   98US-0096822
XX
PA   (PION-) PIONEER HI-BRED INT INC.
XX
PI   Dhugga KS,  Helentjaris TG,  Bowen BA,  Wang X;
XX
DR   WPI; 2000-224343/19.
DR   P-PSDB; AAY84115.
XX
PT   New genes which encode maize cellulose synthase polypeptides in plants
PT   useful for modulating the expression of cellulose synthase in plants
PT   and to produce transgenic plants expressing the novel protein -
XX
PS   Claim 1; Page 144-149; 119pp; English
XX
CC   The present sequence encodes a maize cellulose synthase polypeptide.
CC   The cellulose synthase can be used for the improvement of stalk quality
CC   for improved stand or silage. It also provides an increased concentration
CC   of cellulose in the pericarp, hardening the kernel and improving its
CC   handling ability. The sequences are used to produce transgenic plants
CC   and seeds expressing the cellulose synthase. The polynucleotide is
CC   used for modulating, preferably increasing, the level of the synthase
CC   in a plant cell. The plants are preferably monocots. The polynucleotide
CC   is also used as a probe or primer in the detection, quantitation or
CC   isolation of gene transcripts. The probes are useful in detecting
CC   deficiencies in the level of mRNA in screenings for desired transgenic
CC   plant, for detecting mutations in the gene, for monitoring upregulation
CC   of expression or changes in enzyme activity in screening assays of
CC   compounds, for detection of any number of allelic variants of the gene,
CC   or for use as molecular markers in plant breeding programs. The
CC   isolated nucleic acids of the present invention can also be used for
CC   recombinant expression of their encoded polypeptides or for use as
CC   immunogens in the preparation and/or screening of antibodies. The
CC   proteins can be employed in assays for enzyme agonists or antagonists
CC   of enzyme function or for use of immunogens or antigens to obtain
CC   antibodies specifically immunoreactive with a protein.
XX
SQ   Sequence 3753 BP; 944 A; 805 C; 965 G; 1038 T; 1 other;

Alignment Scores:
Pred. No.:              1.23e-10        Length:         3753
Score:                  162.00          Matches:        23
Percent Similarity:     76.60%          Conservative:   13
Best Local Similarity:  48.94%          Mismatches:     11
Query Match:            50.00%          Indels:         0
DB:                     21              Gaps:           0

US-09-980-043-4 (1-56) x AAZ99515 (1-3753)

Qy          8 IleCysHisThrCysGlyGluGlnValGlyHisAspAlaAsnGlyGluLeuPheValAla 27
              :::|||       ||||||:::   ||||||      |||   |||::::::|||||||||
Db        301 GTCTGCCAGATTTGCGGTGACTCTGTGGGTGTTTCAGCCACTGGTGATGTCTTTGTTGCC 360

Qy         28 CysHisGluCysSerTyrProMetCysLysSerCysPheGluPheGluIleLysGluGly 47
              |||:::||||||::::::|||:::|||:::   |||:::|||:::|||   |||||||||
Db        361 TGCAATGAGTGTGCCTTCCCTGTCTGCCGCCCATGCTATGAGTATGAGCGCAAGGAGGGG 420

Qy         48 ArgLysValCysLeuArgCys 54
                 :::   |||   :::|||
Db        421 AACCAATGCTGCCCCCAGTGC 441

RESULT 12
AAZ99530
ID   AAZ99530 standard; DNA; 3753 BP.
XX
AC   AAZ99530;
XX
DT   03-JUL-2000  (first entry)
XX
DE   DNA encoding a maize cellulose synthase.
XX
KW   Maize; cellulose synthase; stalk quality; stand; silage; cellulose;
KW   transgenic plant; plant breeding marker; ss.
XX
OS   Zea mays
XX
FH   Key             Location/Qualifiers
FT   CDS             184..3405
FT                   /*tag= a
FT                   /product= "cellulose synthase"
FT                   /transl_except= (pos: 916..918, aa: Xaa)
FT                   /note= "no termination codon given; Xaa is an
FT                         unspecified amino acid"
XX
PN   WO200009706-A2.
XX
PD   24-FEB-2000.
XX
PF   16-AUG-1999;   99WO-US18760.
XX
PR   17-AUG-1998;   98US-0096822.
XX
PA   (PION-) PIONEER HI-BRED INT INC.
XX
PI   Dhugga KS,  Helentjaris TG,  Bowen BA,  Wang X;
XX
DR   WPI; 2000-224343/19.
DR   P-PSDB; AAY84120.
```

```
X
T   New genes which encode maize cellulose synthase polypeptides in plants
T   useful for modulating the expression of cellulose synthase in plants
T   and to produce transgenic plants expressing the novel protein -
X
S   Claim 1; Page 184-188; 119pp; English.
X
C   The present sequence encodes a maize cellulose synthase polypeptide.
C   The cellulose synthase can be used for the improvement of stalk quality
C   for improved stand or silage. It also provides an increased concentration
C   of cellulose in the pericarp, hardening the kernel and improving its
C   handling ability. The sequences are used to produce transgenic plants
C   and seeds expressing the cellulose synthase. The polynucleotide is
C   used for modulating, preferably increasing, the level of the synthase
C   in a plant cell. The plants are preferably monocots. The polynucleotide
C   is also used as a probe or primer in the detection quantitation or
C   isolation of gene transcripts. The probes are useful in detecting
C   deficiencies in the level of mRNA in screenings for desired transgenic
C   plant, for detecting mutations in the gene, for monitoring upregulation
C   of expression or changes in enzyme activity in screening assays of
C   compounds, for detection of any number of allelic variants of the gene,
C   or for use as molecular markers in plant breeding programs. The
C   isolated nucleic acids of the present invention can also be used for
C   recombinant expression of their encoded polypeptides or for use as
C   immunogens in the preparation and/or screening of antibodies. The
C   proteins can be employed in assays for enzyme agonists or antagonists
C   of enzyme function or for use of immunogens or antigens to obtain
C   antibodies specifically immunoreactive with a protein.
X
Q   Sequence 3753 BP; 944 A; 805 C; 965 G; 1038 T; 1 other;

lignment Scores:
red. No.:                1.23e-10        Length:          3753
core:                    162.00          Matches:         23
ercent Similarity:       76.60%          Conservative:    13
est Local Similarity:    48.94%          Mismatches:      11
uery Match:              50.00%          Indels:          0
B:                       21              Gaps:            0

S-09-980-043-4 (1-56) x AAZ99530 (1-3753)

Qy          8 IleCysHisThrCysGlyGluGlnValGlyHisAspAlaAsnGlyGluLeuPheValAla 27
              :::|||      ||||||:::   ||||||       |||   |||:::::|||||||||
Db        301 GTCTGCCAGATTTGCGGTGACTCTGTGGGTGTTTCAGCCACTGGTGATGTCTTTGTTGCC 360

Qy         28 CysHisGluCysSerTyrProMetCysLysSerCysPheGluPheGluIleLysGluGly 47
              |||:::||||||:::::::|||:::||:::   |||:::|||:::|||   |||||||||
Db        361 TGCAATGAGTGTGCCTTCCCTGTCTGCCGCCCATGCTATGAGTATGAGCGCAAGGAGGGG 420

Qy         48 ArgLysValCysLeuArgCys 54
                 :::   |||   :::|||
Db        421 AACCAATGCTGCCCCCAGTGC 441

RESULT 13
AAZ99497
 D   AAZ99497 standard; DNA; 3780 BP.
XX
AC   AAZ99497;
XX
DT   03-JUL-2000  (first entry)
XX
DE   DNA encoding a maize cellulose synthase.
XX
KW   Maize; cellulose synthase; stalk quality; stand; silage; cellulose;
KW   transgenic plant; plant breeding marker; ss.
XX
OS   Zea mays.
XX
FH   Key             Location/Qualifiers
FT   CDS             201..3422
FT                   /*tag=  a
FT                   /product= "cellulose synthase"
FT                   /transl_except= (pos: 933..935, aa: Xaa)
FT                   /note  "no termination codon given; Xaa is an
FT                   unspecified amino acid"
XX
PN   WO200009706-A2.
XX
PD   24-FEB-2000.
XX
PF   16-AUG-1999;   99WO-US18760.
XX
PR   17-AUG 1998;   98US-0[illegible]823.
XX
PA   (PION-) PIONEER HI BRED INT INC.
XX
PI   Dhugga KS,  Helentjaris TG,  Powell PA,  Wang X;
XX
DR   WPI; 2000-224343/19.
DR   P-PSDB; AAY84109.
XX
PT   New genes which encode maize cellulose synthase polypeptides in plants
PT   useful for modulating the expression of cellulose synthase in plants
PT   and to produce transgenic plants expressing the novel protein -
XX
PS   Claim 1; Page 87 [illegible]; 87pp; English.
XX
CC   The present sequence encodes a maize cellulose synthase polypeptide.
CC   The cellulose synthase can be used for the improvement of stalk quality
CC   for improved stand or silage. It also provides an increased concentration
CC   of cellulose in the pericarp, hardening the kernel and improving its
CC   handling ability. The sequences are used to produce transgenic plants
CC   and seeds expressing the cellulose synthase. The polynucleotide is
CC   used for modulating, preferably increasing, the level of the synthase
CC   in a plant cell. The plants are preferably monocots. The polynucleotide
CC   is also used as a probe or primer in the detection quantitation or
CC   isolation of gene transcripts. The probes are useful in detecting
CC   deficiencies in the level of mRNA in screenings for desired transgenic
CC   plant, for detecting mutations in the gene, for monitoring upregulation
CC   of expression or changes in enzyme activity in screening assays of
CC   compounds, for detection of any number of allelic variants of the gene,
CC   or for use as molecular markers in plant breeding programs. The
CC   isolated nucleic acids of the present invention can also be used for
CC   recombinant expression of their encoded polypeptides or for use as
CC   immunogens in the preparation and/or screening of antibodies. The
CC   proteins can be employed in assays for enzyme agonists or antagonists
CC   of enzyme function or for use of immunogens or antigens to obtain
CC   antibodies specifically immunoreactive with a protein.
XX
SQ   Sequence 3780 BP; 946 A; 817 C; 976 G; 1040 T; 1 other;

Alignment Scores:
Pred. No.:                1.25e-10        Length:          3780
Score:                    162.00          Matches:         23
Percent Similarity:       76.60%          Conservative:    13
Best Local Similarity:    48.94%          Mismatches:      11
Query Match:              50.00%          Indels:          0
DB:                       21              Gaps:            0

US-09-980-043-4 (1-56) x AAZ99497 (1-3780)

Qy          8 IleCysHisThrCysGlyGluGlnValGlyHisAspAlaAsnGlyGluLeuPheValAla 27
              :::|||      || .||:::   . |||       :||   '||:::::||'|||||'
Db        318 GTCTGCCAGATTT[illegible]GGTGACTCTGTGGGTGTTTCAGCCACTGGTGATGTCTTTGTTGCC 377

Qy         28 CysHisGluCysSerTyrProMetCysLysSerCysPheGluPheGluIleLysGluGly 47
              |||:::|||  .|:::::::||:::.||:::   .||:::||'::||    ||'||||.|
Db        378 TGCAATGAGTGTGC[illegible]TTCCCTGTCTGCCGCCCATGCTATGAGTATGAGCGCAAGGAGGGG 437

Qy         48 ArgLysValCysLeuArgCys 54
                 :::   |||   :::|:
Db        438 AACCAATGCTGCCCCCAGTGC 458

RESULT 14
AAV06569
ID   AAV06569 standard; cDNA; 1741 BP.
```

```
XX
AC   AAV06569;
XX
DT   06-JUL-1998  (first entry)
XX
DE   Rice cellulose synthase RSW1 like cDNA clone S0543.
XX
KW   Cellulose synthase; RSW1-like gene; beta-1,4-glucan;
KW   transgenic plant; rice; ss.
XX
OS   Oryza sativa.
XX
FH   Key             Location/Qualifiers
FT   CDS             101..1741
FT                   /*tag= a
XX
PN   WO9800549-A1.
XX
PD   08-JAN-1998.
XX
PF   24-JUN-1997;   97WO-AU00402.
XX
PR   27-JUN-1996;   96AU-0000699.
XX
PA   (CSIR ) COMMONWEALTH SCI & IND RES ORG.
PA   (AUSU ) UNIV AUSTRALIAN NAT.
XX
PI   Arioli A,  Betzner AS,  Peng L,  Williamson RE;
XX
DR   WPI; 1998-086974/08.
DR   P-PSDB; AAW46202
XX
PT   DNA encoding cellulose biosynthetic enzyme - useful for manipulation
PT   of cellulose and beta-1,4-glucan
XX
PS   Claim 13; Page 163-166; 207pp; English.
XX
CC   Rice partial cDNA clone EST S0543 was isolated on the basis of
CC   homology to newly identified Arabidopsis cellulose synthase RSW1
CC   nucleotide sequences (see AAV06562 68).  It encodes the first 547
CC   amino acids (see AAW46202) of the rice RSW1 or RSW1-like polypeptide,
CC   showing 70% identity to the Arabidopsis RSW1 polypeptide given in
CC   AAW33817.  Claimed nucleic acid molecules (see AAV06562 69) coding for
CC   claimed polypeptides (see AAW33816 20 and AAW46202) involved in
CC   cellulose biosynthesis can be used to manipulate the cellulose
CC   and/or beta-glucan content of transgenic plants.  Expression of
CC   nucleic acids in the sense orientation increases the level of
CC   cellulose and reduces the level of non crystalline beta 1,4 glucan
CC   and starch, providing plants with modified strength and/or shape
CC   and/or fibre properties, or having increased resistance to stresses
CC   or pests.  Antisense, ribozyme or co suppression molecules can be
CC   used to reduce the cellulose content of a transgenic plant, e.g. to
CC   improve digestibility or to alter carbon partitioning such that
CC   increased carbon is available for growth.
XX
SQ   Sequence 1741 BP; 440 A; 399 C; 486 G; 416 T; 0 other;

Alignment Scores:
Pred. No.:              6.2e-11       Length:          1741
Score:                  161.00        Matches:         23
Percent Similarity:     76.60%        Conservative:    13
Best Local Similarity:  48.94%        Mismatches:      11
Query Match:            49.69%        Indels:          0
DB:                     19            Gaps:            0

US-09-980-043-4 (1-56) x AAV06569 (1-1741)

Qy          8 IleCysHisThrCysGlyGluGlnValGlyHisAspAlaAsnGlyGluLeuPheValAla 27
              :::| |      |||||:::   ||||||     |||   |||:::::|||||||:|
Db        215 GTCTGCCAGATTTGTGGCGACACTGTTGGCGTCTCGGCCACCGGCGACGTCTTTGTTGCC 277

Qy         28 CysHisGluCysSerTyrProMetCysLysSerCysPheGluPheGluIleLysGluGly 47
              |:|:::|||||:|::::::|||:::|||:::    |||:::|:::||   |||||||:|

Db        278 TGCAATGAGTGCGCCTTCCCGGTCTGCCGCCCTTGCTACGAGTACGAACGCAAGGAAGGG 337

Qy         48 ArgLysValCysLeuArgCys 54
                 :::   |||   :::| |
Db        338 AACCAGTGCTGCCCCCAGTGC 358

RESULT 15
AAZ99500
ID   AAZ99500 standard; DNA; 3725 BP.
XX
AC   AAZ99500;
XX
DT   03-JUL-2000  (first entry)
XX
DE   DNA encoding a maize cellulose synthase.
XX
KW   Maize; cellulose synthase; stalk quality; stand; silage; cellulose;
KW   transgenic plant; plant breeding marker; ss.
XX
OS   Zea mays.
XX
FH   Key             Location/Qualifiers
FT   CDS             179..3398
FT                   /*tag= a
FT                   /product= "cellulose synthase"
FT                   /note= "no termination codon given"
XX
PN   WO200009706-A2.
XX
PD   24-FEB-2000.
XX
PF   16-AUG-1999;   99WO-US18760.
XX
PR   17-AUG-1998;   98US-0096822.
XX
PA   (PION-) PIONEER HI-BRED INT INC.
XX
PI   Dhugga KS,  Helentjaris TG,  Bowen BA,  Wang X;
XX
DR   WPI; 2000-224343/19.
DR   P-PSDB; AAY84110.
XX
PT   New genes which encode maize cellulose synthase polypeptides in plants
PT   useful for modulating the expression of cellulose synthase in plants
PT   and to produce transgenic plants expressing the novel protein -
XX
PS   Claim 1; Page 105-110; 119pp; English.
XX
CC   The present sequence encodes a maize cellulose synthase polypeptide.
CC   The cellulose synthase can be used for the improvement of stalk quality
CC   for improved stand or silage. It also provides an increased concentration
CC   of cellulose in the pericarp, hardening the kernel and improving its
CC   handling ability. The sequences are used to produce transgenic plants
CC   and seeds expressing the cellulose synthase. The polynucleotide is
CC   used for modulating, preferably increasing, the level of the synthase
CC   in a plant cell. The plants are preferably monocots. The polynucleotide
CC   is also used as a probe or primer in the detection quantitation or
CC   isolation of gene transcripts. The probes are useful in detecting
CC   deficiencies in the level of mRNA in screenings for desired transgenic
CC   plant, for detecting mutations in the gene, for monitoring upregulation
CC   of expression or changes in enzyme activity in screening assays of
CC   compounds, for detection of any number of allelic variants of the gene,
CC   or for use as molecular markers in plant breeding programs. The
CC   isolated nucleic acids of the present invention can also be used for
CC   recombinant expression of their encoded polypeptides or for use as
CC   immunogens in the preparation and/or screening of antibodies. The
CC   proteins can be employed in assays for enzyme agonists or antagonists
CC   of enzyme function or for use of immunogens or antigens to obtain
CC   antibodies specifically immunoreactive with a protein.
XX
SQ   Sequence 3725 BP; 927 A; 807 C; 950 G; 1041 T; 0 other;

Alignment Scores:
```

red. No.:	1.65e-10	Length:	3725
core:	161.00	Matches:	23
ercent Similarity:	76.60%	Conservative:	13
est Local Similarity:	48.94%	Mismatches:	11
uery Match:	49.69%	Indels:	0
B:	21	Gaps:	0

S-09-980-043-4 (1-56) x AAZ99500 (1-3725)

```
y          8 IleCysHisThrCysGlyGluGlnValGlyHisAspAlaAsnGlyGluLeuPheValAla 27
             :::|||       ||||||:::   ||||||       |||   |||::::::|||| ||||
b        296 GTCTGCCAGATTTGTGGCGACACTGTTGGCGTTTCAGCCACTGGTGATGTCTTTGTTGCC 355

y         28 CysHisGluCysSerTyrProMetCysLysSerCysPheGluPheGluIleLysGluGly 47
             |||:::||||||:::::::|||:::|||:::   |||:::|||:::|||   ||||||||:
b        356 TGCAATGAGTGTGCCTTCCCTGTCTGCCGCCCTTGCTATGAGTACGAGCGCAAGGAAGGG 415

y         48 ArgLysValCysLeuArgCys 54
                :::   |||   :::|||
b        416 AACCAATGCTGCCCTCAGTGC 436
```

earch completed: October 31, 2003, 16:00:31
ob time : 43.2147 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 31, 2003, 15:18:06 ; Search time 602.731 Seconds
(without alignments)
3800.930 Million cell updates/sec

Title: US-09-980-043-4
Perfect score: 324
Sequence: 1 MMESGAPICHTCGEQVGHDA.....KSCFEFEIKEGRKVCLRCGS 56

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09980043/runat_29102003_163852_21863/app_query.fasta_1.1358
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=USC9980043_@CGN_1_1_6944_@runat_29102003_163852_21863 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	324	100.0	1010	8	AF197911	AF197911 Populus t
2	319	98.5	3232	8	AF072131	AF072131 Populus t
3	264	81.5	3186	8	GHU58283	U58283 Gossypium h
4	264	81.5	3328	6	AR164434	AR164434 Sequence
5	264	81.5	6886	8	AF413210	AF413210 Gossypium
6	247	76.2	2958	8	AF267742	AF267742 Arabidops
7	246	75.9	3109	8	AY099636	AY099636 Arabidops
c 8	246	75.9	94301	8	ATF28A21	AL035526 Arabidops
c 9	246	75.9	199075	8	ATCHRIV49	AL161549 Arabidops
10	220	67.9	1095	8	AF139442	AF139442 Gossypium
11	220	67.9	1095	8	AF139443	AF139443 Gossypium
12	220	67.9	1096	8	AF139446	AF139446 Gossypium
13	220	67.9	1097	8	AF139444	AF139444 Gossypium
14	211	65.1	1097	8	AF139445	AF139445 Gossypium
15	211	65.1	1201	8	AF139449	AF139449 Gossypium
16	211	65.1	1202	8	AF139451	AF139451 Gossypium
17	202	62.3	1201	8	AF139450	AF139450 Gossypium
18	201	62.0	1207	8	AF139447	AF139447 Gossypium
19	201	62.0	1207	8	AF139448	AF139448 Gossypium
20	193.5	59.7	3145	8	AF323039	AF323039 Zinnia el
21	192	59.3	1076	8	AY125066	AY125066 Gossypium
22	192	59.3	1076	8	AY125067	AY125067 Gossypium
23	192	59.3	1076	8	AY125068	AY125068 Gossypium
24	192	59.3	1076	8	AY125069	AY125069 Gossypium
25	192	59.3	1077	8	AF419971	AF419971 Gossypium
26	191	59.0	1076	8	AY125065	AY125065 Gossypium
27	189	58.3	1072	8	AF201886	AF201886 Gossypioi
28	189	58.3	1072	8	AF419973	AF419973 Gossypium
29	189	58.3	1076	8	AF419972	AF419972 Gossypium
30	188	58.0	1076	8	AF419970	AF419970 Gossypium
31	187	57.7	1009	8	AF376047	AF376047 Cienfuego
32	187	57.7	1058	8	AF376042	AF376042 Hampea ap
33	187	57.7	1064	8	AF376043	AF376043 Thespesia
34	187	57.7	1064	8	AF376044	AF376044 Thespesia
35	184	56.8	1055	8	AF376045	AF376045 Thespesia
36	183	56.5	1064	8	AF376046	AF376046 Lebronnec
37	183	56.5	1075	8	AF376040	AF376040 Kokia dry
38	183	56.5	137056	2	AC131248	AC131248 Medicago
39	181	55.9	996	8	AF376048	AF376048 Malva syl
40	176	54.3	1116	8	AF419978	AF419978 Kokia dry
41	174	53.7	1158	8	AY125072	AY125072 Gossypium
42	174	53.7	1169	8	AY125071	AY125071 Gossypium
43	174	53.7	1170	8	AF419977	AF419977 Gossypium
44	174	53.7	1170	8	AY125070	AY125070 Gossypium
45	174	53.7	1170	8	AY125074	AY125074 Gossypium

ALIGNMENTS

RESULT 1

```
F197911
OCUS       AF197911               1010 bp    DNA     linear   PLN 02-AUG-2000
EFINITION  Populus tremuloides cellulose synthase (CesA) gene, promotor and
           partial cds.
CCESSION   AF197911
ERSION     AF197911.1  GI:9651405
EYWORDS    .
OURCE      Populus tremuloides (quaking aspen)
 ORGANISM  Populus tremuloides
           Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
           Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
           rosids; eurosids I; Malpighiales; Salicaceae; Populus.
EFERENCE   1  (bases 1 to 1010)
 AUTHORS   Wu,L., Joshi,C.P. and Chiang,V.L.
 TITLE     A xylem-specific cellulose synthase gene from aspen (Populus
           tremuloides) is responsive to mechanical stress
 JOURNAL   Plant J. 22 (6), 495-502 (2000)
 MEDLINE   20344632
  PUBMED   10886769
EFERENCE   2  (bases 1 to 1010)
 AUTHORS   Wu,L., Joshi,C.P. and Chiang,V.L.
 TITLE     Direct Submission
 JOURNAL   Submitted (22-OCT-1999) Forestry, Michigan Tech University, 1400,
           Townsend Drive, Houghton, MI 49931, USA
EATURES             Location/Qualifiers
     source         1. .1010
                    /organism="Populus tremuloides"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:3693"
                    /tissue_type="xylem"
     gene           <1. .>1010
                    /gene="CesA"
     misc_feature   1. .840
                    /gene="CesA"
                    /note="contains promotor and 5'UTR"
     mRNA           <841. .>1010
                    /gene="CesA"
                    /product="cellulose synthase"
     CDS            841. .>1010
                    /gene="CesA"
                    /codon_start=1
                    /product="cellulose synthase"
                    /protein_id="AAF91155.1"
                    /db_xref="GI:9651406"
                    /translation="MMESGAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEI
                    KEGRKVCLRCGS"
ASE COUNT      289 a    212 c    216 g    293 t
RIGIN
```

```
Alignment Scores:
Pred. No.:                8.3e-30        Length:          1010
Score:                    324.00         Matches:         56
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       8              Gaps:            0

US-09-980-043-4 (1-56) x AF197911 (1-1010)

Qy          1 MetMetGluSerGlyAlaProIleCysHisThrCysGlyGluGlnValGlyHisAspAla 20
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 ATGATGGAATCTGGGGCTCCTATATGCCATACCTGTGGTGAACAGGTGGGGCATGATGCA 900

Qy         21 AsnGlyGluLeuPheValAlaCysHisGluCysSerTyrProMetCysLysSerCysPhe 40
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AATGGGGAGCTATTTGTGGCTTGCCATGAGTGTAGCTATCCCATGTGCAAGTCTTGTTTC 960

Qy         41 GluPheGluIleLysGluGlyArgLysValCysLeuArgCysGlySer 56
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GAGTTTGAAATCAAAGAGGGCCGGAAAGTTTGCTTGCGGTGTGGCTCG 1008

RESULT 2
AF072131
```

```
LOCUS       AF072131               3232 bp    mRNA    linear   PLN 26-APR-2001
DEFINITION  Populus tremuloides secondary xylem cellulose synthase (celA) mRNA,
            complete cds.
ACCESSION   AF072131
VERSION     AF072131.1  GI:4115904
KEYWORDS    .
SOURCE      Populus tremuloides (quaking aspen)
  ORGANISM  Populus tremuloides
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Malpighiales; Salicaceae; Populus
REFERENCE   1  (bases 1 to 3232)
  AUTHORS   Wu,L., Joshi,C.P. and Chiang,V.L.
  TITLE     A xylem-specific cellulose synthase gene from aspen (Populus
            tremuloides) is responsive to mechanical stress
  JOURNAL   Plant J. 22 (6), 495-502 (2000)
  MEDLINE   20344632
   PUBMED   10886769
REFERENCE   2  (bases 1 to 3232)
  AUTHORS   Wu,L., Joshi,C.P. and Chiang,V.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JUN-1998) IWR/Forestry, Michigan Tech University,
            1400, Townsend Drive, Houghton, MI 49931, USA
REFERENCE   3  (bases 1 to 3232)
  AUTHORS   Wu,L., Joshi,C.P. and Chiang,V.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JAN-1999) IWR/Forestry, Michigan Tech University,
            1400, Townsend Drive, Houghton, MI 49931, USA
  REMARK    Sequence update by submitter
COMMENT     On Jan 8, 1999 this sequence version replaced gi:3243277.
FEATURES             Location/Qualifiers
     source          1. .3232
                     /organism="Populus tremuloides"
                     /mol_type="mRNA"
                     /db_xref="taxon:3693"
                     /tissue_type="secondary xylem"
     gene            1. .3232
                     /gene="celA"
     CDS             69. .3005
                     /gene="celA"
                     /codon_start=1
                     /product="secondary xylem cellulose synthase"
                     /protein_id="AAD03417.1"
                     /db_xref="GI:4115905"
                     /translation="MMESGAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEI
                     NEGRKVCLRCGSPYDENLLDDVEKKGSGNQSTMASHLNDSQDVGIHARHISSVSTVDS
                     EMNDEYGNPIWKNRVKSCKDKENKKKKRSPKAETEPAQVPTEQQMEEKPSAEASEPLS
                     IVYPIPRNKLTPYRAVIIMRLVILGLFFHFRITNPVDSAFGLWLTSVICEIWFAFSWV
                     LDQFPKWNPVNRETYIERLSARYEREGEPSQLAGVDFFVSTVDPLKEPPLITANTVLS
                     ILAVDYPVDKVSCYVSDDGAAMLSFESLVETAEFARKWVPFCKKFSIEPRAPEFYFSQ
                     KIDYLKDKVQPSFVKERRAMKRDYEEYKVRVNALVAKAQKTPEEGWTMQDGTPWPGNN
                     TRDHPGIIQVFLGHSGGLDPWEILGARDIEGNELPRLVYVSREKRPGYQHHKKAGAENALVRVSAV
                     LTNAPYILNVDCDHYVNNSKAVREAMCILMDPQVGRDVCYVQFPQRFDGIDKSDRYAN
                     RNVVFFDVNMKGLDGIQGPVYVGTGCVFNRQALYGYGPPSMPSLRKRKDSSSCFSCCC
                     PSKKKPAQDPAEVYRDAKREDLNAAIFNLTEIDNYDEHERSMLISQLSFEKTFGLSSV
                     FIESTLMENGGVPESANSPPFIKEAIQVIGCGYEEKTEWGKQIGWIYGSVTEDILSGF
                     KMHCRGWRSIYCMPVRPAFKGSAPINLSDRLHQVLRWALGSVEIFFSRHCPLWYGFGG
                     GRLKWLQRLAYINTIVYPFTSLPLIAYCTIPAVCLLTGKFIIPTLSNLASMLFLGLFI
                     SIIVTAVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQGFLKMLAGIDTNFTVTA
                     KAAEDAEFGELYMVKWTTLLIPPTTLLIINMSGCAGFSDALNKGYEAWGPLFGKVFFA
                     FWVILHLYPFLKGLMGRQNLTPTIVVLWSVLLASVFSLVWVKINPFVNKVDNTLVAET
                     CISIDC"
BASE COUNT      895 a    668 c    772 g    897 t
ORIGIN
```

```
Alignment Scores:
Pred. No.:                1.14e-28       Length:          3232
Score:                    319.00         Matches:         55
Percent Similarity:       98.21%         Conservative:    0
Best Local Similarity:    98.21%         Mismatches:      1
Query Match:              98.46%         Indels:          0
DB:                       8              Gaps:            0
```

```
US-09-980-043-4 (1-56) x AF072131 (1-3232)

Qy          1 MetMetGluSerGlyAlaProIleCysHisThrCysGlyGluGlnValGlyHisAspAla 20
              ||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||!
Db         69 ATGATGGAATCTGGGGCTCCTATATGCCATACCTGTGGTGAACAGGTGGGGCATGATGCA 128

Qy         21 AsnGlyGluLeuPheValAlaCysHisGluCysSerTyrProMetCysLysSerCysPhe 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||!
Db        129 AATGGGGAGCTATTTGTGGCTTGCCATGAGTGTAGCTATCCCATGTGCAAGTCTTGTTTC 188

Qy         41 GluPheGluIleLysGluGlyArgLysValCysLeuArgCysGlySer 56
              |||||||||||   ||||||| ||||| ||||||||||||||||||||
Db        189 GAGTTTGAAATCAATGAGGGCCGCAAAGTTGCTTGCGGTGTGGCTCG 236

RESULT 3
GHU58283
LOCUS       GHU58283                3186 bp    mRNA    linear   PLN 05-DEC-1996
DEFINITION  Gossypium hirsutum cellulose synthase (celA1) mRNA, complete cds.
ACCESSION   U58283
VERSION     U58283.1  GI:1706955
KEYWORDS    .
SOURCE      Gossypium hirsutum (upland cotton)
  ORGANISM  Gossypium hirsutum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium.
REFERENCE   1  (bases 1 to 3186)
  AUTHORS   Pear,J.R., Kawagoe,Y., Schreckengost,W.E., Delmer,D.P. and
            Stalker,D.M.
  TITLE     Higher plants contain homologs of the bacterial celA genes encoding
            the catalytic subunit of cellulose synthase
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 93 (22), 12637-12642 (1996)
  MEDLINE   97057296
   PUBMED   8901635
REFERENCE   2  (bases 1 to 3186)
  AUTHORS   Pear,J.R., Kawagoe,Y., Schreckengost,W.E., Delmer,D.P. and
            Stalker,D.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-MAY-1996) J.R. Pear, Calgene, Inc, 1920 5th Street,
            Davis, CA 95616, USA
FEATURES             Location/Qualifiers
     source          1. .3186
                     /organism="Gossypium hirsutum"
                     /mol_type="mRNA"
                     /strain="Acala SJ-2"
                     /db_xref="taxon:3635"
                     /tissue_type="fiber"
                     /dev_stage="17-21 days post anthesis"
     gene            1. .3186
                     /gene="celA1"
     CDS             77. .3001
                     /gene="celA1"
                     /codon_start=1
                     /product="cellulose synthase"
                     /protein_id="AAB37766.1"
                     /db_xref="GI:1706956"
                     /translation="MMESGVPVCHTCGEHVGLNVNGEPFVACHECNFPICKSCFEYDL
                     KEGRKACLRCGSPYDENLLDDVEKATGDQSTMAAHCNKSQDVGIHARHISSVSTLDSE
                     MAEDNGNSIWKNRVESWKEKKNKKKKPATTKVERPAEIPPEQQMEDKPAPDASQPLST
                     IIPIPKSRLAPYRTVIIMRLIILGLFFHYRVTNPVDSAFGLWLTSVICEIWFAFSWVL
                     DQFPKWYPVNRETYIDRLSARYEREGEPDELAAVDFFVSTVDPLKEPPLITANTVLSI
                     LALDYPVDKVSCYISDDGAAMLTFESLVETADFARKWVPFCKKFSIEPRAPEFYFSQK
                     IDYLKDKVQPSFVKERRAMKRDYEEYKIRINALVAKAQKTPDEGWTMQDGTSWPGNNP
                     RDHPGMIQVFLGYSGARDIEGNELPRLVYVSREKRPGYQHHKKAGAENALVRVSAVLT
                     NAPFILNLDCDHYVNNSKAVREAMCFLMDPQVGRDVCYVQFPQRFDGIDRSDRYANRN
                     TVFFDVNMKGLDGIQGPVYVGTGCVFNRQALYGYGPPSMPSFPKSSSSSCSCCCPGKK
                     EPKDPSELYRDAKREELDAAIFNLREIDNYDEYERSMLISQTSFEKTFGLSSVFIEST
                     LMENGGVAESANPSTLIKEAIHVISCGYEEKTAWGKEIGWIYGSVTEDILTGFKMHCR
                     GWRSIYCMPLRPAFKGSAPINLSDRLHQVLRWALGSVEIFLSRHCPLWYGFGGRLKW
                     LQRLAYINTIVYPFTSLPLIAYCSLPAICLLTGKFIIPTLSNLASVLFLGLFLSIIVT
                     AVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQGFLKMLAGIDTNFTVTAKAADD
                     ADFGELYIVKWTTLLIPPTTLLIVNMVGVVAGFSDALNKGYEAWGPLFGKVFFSFWVI
                     LHLYPFLKGLMGRQNRTPTIVVLWSVLLASVFSLLWVRINPFVSTADSTTVEMSLDFI

                     DC"
BASE COUNT      864 a    654 c    723 g    945 t
ORIGIN

Alignment Scores:
Pred. No.:                4.83e-22       Length:         3186
Score:                    264.00         Matches:        42
Percent Similarity:       89.29%         Conservative:   8
Best Local Similarity:    75.00%         Mismatches:     6
Query Match:              81.48%         Indels:         0
DB:                       8              Gaps:           0

US-09-980-043-4 (1-56) x GHU58283 (1-3186)

Qy          1 MetMetGluSerGlyAlaProIleCysHisThrCysGlyGluGlnValGlyHisAspAla 20
              |||||||||||||||||   |||:::|||||||||||||||||||   ||||||   :::
Db         77 ATGATGGAATCTGGGGTTCCTGTTTGCCACACTTGTGGTGAACATGTTGGGTTGAATGTT 136

Qy         21 AsnGlyGluLeuPheValAlaCysHisGluCysSerTyrProMetCysLysSerCysPhe 40
              |||||||||   ||||||||||||||||||||||:::::|||:::|||||||||||||||
Db        137 AATGGTGAACCTTTTGTGGCTTGCCATGAATGTAATTTCCCTATTTGTAAGAGTTGTTTT 196

Qy         41 GluPheGluIleLysGluGlyArgLysValCysLeuArgCysGlySer 56
              |||:::::::::|||||||||||||||   ||||||||||||||||||
Db        197 GAGTATGATCTTAAGGAAGGACGAAAAGCTTGCTTGCGTTGTGGTAGT 244

RESULT 4
AR164434
LOCUS       AR164434                3328 bp    DNA     linear   PAT 17-OCT-2001
DEFINITION  Sequence 1 from patent US 6271443.
ACCESSION   AR164434
VERSION     AR164434.1  GI:16235654
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 3328)
  AUTHORS   Stalker,D.M., Pear,J.R. and Delmer,D.
  TITLE     Cotton and rice cellulose synthase DNA sequences
  JOURNAL   Patent: US 6271443-A 1 07-AUG-2001;
FEATURES             Location/Qualifiers
     source          1. .3328
                     /organism="unknown"
BASE COUNT      914 a    689 c    763 g    962 t
ORIGIN

Alignment Scores:
Pred. No.:                5.06e-22       Length:         3328
Score:                    264.00         Matches:        42
Percent Similarity:       89.29%         Conservative:   8
Best Local Similarity:    75.00%         Mismatches:     6
Query Match:              81.48%         Indels:         0
DB:                       6              Gaps:           0

US-09-980-043-4 (1-56) x AR164434 (1-3328)

Qy          1 MetMetGluSerGlyAlaProIleCysHisThrCysGlyGluGlnValGlyHisAspAla 20
              |||||||||||||||||   |||:::|||||||||||||||||||   ||||||   :::
Db        179 ATGATGGAATCTGGGGTTCCTGTTTGCCACACTTGTGGTGAACATGTTGGGTTGAATGTT 238

Qy         21 AsnGlyGluLeuPheValAlaCysHisGluCysSerTyrProMetCysLysSerCysPhe 40
              |||||||||   ||||||||||||||||||||||:::::|||:::|||||||||||||||
Db        239 AATGGTGAACCTTTTGTGGCTTGCCATGAATGTAATTTCCCTATTTGTAAGAGTTGTTTT 298

Qy         41 GluPheGluIleLysGluGlyArgLysValCysLeuArgCysGlySer 56
              |||:::::::::|||||||||||||||   ||||||||||||||||||
Db        299 GAGTATGATCTTAAGGAAGGACGAAAAGCTTGCTTGCGTTGTGGTAGT 346

RESULT 5
AF413210
LOCUS       AF413210                6886 bp    DNA     linear   PLN 27-DEC-2001
DEFINITION  Gossypium hirsutum cellulose synthase A4 (CesA4) gene, complete
            cds.
```

```
CCESSION   AF413210
ERSION     AF413210.1  GI:17226293
EYWORDS    .
OURCE      Gossypium hirsutum (upland cotton)
 ORGANISM  Gossypium hirsutum
           Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
           Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
           rosids; eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium.
EFERENCE   1  (bases 1 to 6886)
 AUTHORS   Kim,H.J. and Triplett,B.A.
 TITLE     Cotton fiber growth in planta and in vitro. Models for plant cell
           elongation and cell wall biogenesis
 JOURNAL   Plant Physiol. 127 (4), 1361-1366 (2001)
 MEDLINE   21608363
  PUBMED   11743074
EFERENCE   2  (bases 1 to 6886)
 AUTHORS   Kim,H.J., Williams,M.Y. and Triplett,B.A.
 TITLE     Direct Submission
 JOURNAL   Submitted (22-AUG-2001) Fiber Bioscience, USDA-ARS, SRRC, 1100
           Robert E. Lee Blvd, New Orleans, LA 70124, USA
EATURES             Location/Qualifiers
     source         1. .6886
                    /organism="Gossypium hirsutum"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:3635"
     gene           <2685. .>6766
                    /gene="CesA4"
     mRNA           join(<2685. .2862,3033. .3106,3193. .3244,3338. .3482,
                    3613. .3879,3992. .4337,4426. .4689,4778. .4990,5085. .5328,
                    5420. .5619,5732. .6085,6179. .>6766)
                    /gene="CesA4"
                    /product="cellulose synthase A4"
     CDS            join(2685. .2862,3033. .3106,3193. .3244,3338. .3482,
                    3613. .3879,3992. .4337,4426. .4689,4778. .4990,5085. .5328,
                    5420. .5619,5732. .6085,6179. .6766)
                    /gene="CesA4"
                    /codon_start=1
                    /product="cellulose synthase A4"
                    /protein_id="AAL37718.1"
                    /db_xref="GI:17226294"
                    /translation="MMESGVPVCHTCGEHVGLNVNGEPFVACHECNFPICKSCFEYDL
                    KEGRKACLRCGSPYDENLLDDVEKTTGDQSTMAAHLSKSQDVGIHARHISSVSTLDSE
                    MTGDNGNPIWKNRVESWKEKKNKKKKPATTKVEREAEIPPEQQMEDKPAPDASQPLST
                    IIPIPKSRLAPYRTVIIMRLIILGLFFHYRVTHPVDSAFGLWLTSVICEIWFAFSWVL
                    DQFPKWYPVNRETYIDRLSARYDREGEPNELAAVDFFVSTVDPLKEPPLITANTVLSI
                    LALDYPVDKVSCYISDDGAAMLTFESLVETADFARKWVPFCKKFSIEPRAPEFYFSQK
                    IVYLKDKVQPSFVKERRAMKRDYEEYKIRINALVAKAQKTPEEGWTMQDGTPWPGNNP
                    RDHPGMIQVFLGYSGAHDIEGNELPRLVYVSREKRPGYQHHKQAGADNALVRVSAALT
                    NAPPILNLDCDHYVNNSKAVREAMCCLMDPQGGRDVCYVQFPQRCDGIDRSERYAKRN
                    TVFFDVNMKGRDGSQGPVYVGTGCVCNRQALYGYGPPSMPSFPKSSSSSCSCCCPGKK
                    EPKEPTELYRDAKREELDAAIFNLREIDNYDEYERSMLISQTSFEKTFGLSSVFIEST
                    LMENGGVAESANPSTLIKEAIHVIGCGYEEKTAWGKEIGWIYGSVTEDILTGFKMHCR
                    GWRSIYCMPLRPAFKGSAPINLSDRLHQVLRWALGSLEIFLSRHCPLWYGFGGGRLKW
                    LQRLAYINTSVYPFTSLPLIAYCSLPAICLLTGKFIIPTLSNLASVLYLGLFLSIIVT
                    AVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQGFLKMLAGIDTNFTVTAKAADD
                    ADFGELYIVKWTTLLIPPTTLLIVNMVGVVAGFSDALNKGYEAWGPLFGKVFFSFWVI
                    LHLYPFLKGLMGRQNRTPTIVVLWSVLLASVFSLVWVRINPFVSTADSTTVSRANSSI
                    DC"
ASE COUNT      2067 a    1237 c    1277 g    2304 t       1 others
RIGIN

Alignment Scores:
Pred. No.:                1.09e-21        Length:          6886
Score:                    264.00          Matches:         42
Percent Similarity:       89.29%          Conservative:    8
Best Local Similarity:    75.00%          Mismatches:      6
Query Match:              81.48%          Indels:          0
DB:                       8               Gaps:            0

US-09-980-043-4 (1-56) x AF413210 (1-6886)

Qy          1 MetMetGluSerGlyAlaProIleCysHisThrCysGlyGluGlnValGlyHisAspAla 20
              ||||||||||||||||   |||:::|||||||||||||||||||   ||||||   :::

Db       2685 ATGATGGAATCTGGGGTTCCTGTTTGTGTAGA TTGTGGTGAACATGTTGGGTTGAATGTT 2744

Qy         21 AsnGlyGluLeuProValAlaCysHisGluCysSerTyrProMetCysLysSerCysPhe 40
              |||:||||:  ||, |||| ||| ::||| .:::::|||:::|||||| |||| |.|

Db       2745 AATGGTGAACCCTTTGTGGCTTGCCATGAATGTAATTTCCCTATTTGTAAGAGTTGTTTT 2804

Qy         41 GluPheGluIleLysGluGlyArgLysValCysLeuArgCysGlySer 56
              |||::::::::||:|||||.|||'.  |:.||||||||||||||||
Db       2805 GAGTATGATCTTAAGGAAGGACGAAAAGCTTGCTTGCGTTGTGGTAGT 2852

RESULT 6
AF267742
LOCUS       AF267742                2958 bp    mRNA    linear   PLN 16 FEB 2001
DEFINITION  Arabidopsis thaliana cellulose synthase catalytic subunit (IRX1)
            mRNA, complete cds.
ACCESSION   AF267742
VERSION     AF267742.1  GI:12836996
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis
REFERENCE   1  (bases 1 to 2958)
  AUTHORS   Taylor,N.G., Laurie,S. and Turner,S.R.
  TITLE     Multiple cellulose synthase catalytic subunits are required for
            cellulose synthesis in Arabidopsis
  JOURNAL   Plant Cell 12 (12), 2529-2540 (2000)
  MEDLINE   21066939
   PUBMED   11148295
REFERENCE   2  (bases 1 to 2958)
  AUTHORS   Taylor,N.G. and Turner,S.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAY-2000) Biological Sciences, University of
            Manchester, 3.614 Stopford Building, Oxford Road, Manchester M13
            9PT, UK
FEATURES             Location/Qualifiers
     source          1. .2958
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
     gene            1. .2958
                     /gene="IRX1"
     CDS             1. .2958
                     /gene="IRX1"
                     /codon_start=1
                     /product="cellulose synthase catalytic subunit"
                     /protein_id="AAK08700.1"
                     /db_xref="GI:12836997"
                     /translation="MMESRSPICNTCGEEIGAKSNGEFFVACHECSFPICKACLEYEF
                     KEGRRICLRCGNPYDENVFDDVETKTSKTQSIVPTQTNNTSQDSGIHARHISTVSTID
                     SELNDEYGNPIWKNRVESWKDKKDKKSKKKKKDPKATKAEQQDAQVPTQQHMEDMPPN
                     TESGATDVLSVVIPIPRTKITSYRIVIIMRLIILALFFNYRITHPVDSAYGLWLTSVI
                     CEIWFAVSWVLDQFPKWSPINRETYIDRLSARFEREGEQSQLAAVDFFVSTVDPLKEP
                     PLITANTVLSILALDYPVDKVSCYVSDDGAAMLSFESLVETADFARKWVPFCKKYSIE
                     PRAPEFYFSLKIDYLRDKVQPSFVKERRAMKRDYEEFKIRMNALVAKAQKTPEEGWTM
                     QDGTSWPGNNTRDHPGMIQVFLGYSGARDIEGNELPRLVYVSREKRPGYQHHKKAGAE
                     NALVRVSAVLTNAPFILNLDCDHYVNNSKAVREAMCFLMDPVVGQDVCFVQFPQRFDG
                     IDKSDRYANRNIVFFDVNMRGLDGIQGPVYVGTGTVFRRQALYGYSPPSKPRILPQSS
                     SSSCCCLTKKKQPQDPAEIYKDAKREELDAAIFNLGDLDNYDEYDRSMLISQTSFEKT
                     FGLSTVFIESTLMENGGVPDSVNPSTLIKEAIHVISCGYEEKTEWGKEIGWIYGSITE
                     DILTGFKMHCRGWRSIYCMPLRPAFKGSAPINLSDRLHQVLRWALGSVEIFLSRHCPL
                     WYGCSGGRLKLLQRLAYINTIVYPFTSLPLVAYCTLPAICLLTGKFIIPTLSNLASML
                     FLGLFISIILTSVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQGFLKMLAGLDT
                     NFTVTSKTADDLEFGELYIVKWTTLLIPPTSLLIINLVGVVAGFSDALNKGYEAWGPL
                     FGKVFFAFWVILHLYPFLKGLMGRQNRTPTIVILWSILLASVFSLVWVRINPFVSKTD
                     TTSLSLNCLLIDC"
BASE COUNT      818 a    613 c    658 g    872 t
ORIGIN

Alignment Scores:
Pred. No.:                5.02e-20        Length:          2958
Score:                    247.00          Matches:         38
```

```
Percent Similarity:     89.29%          Conservative:  12
Best Local Similarity:  67.86%          Mismatches:    6
Query Match:            76.23%          Indels:        0
DB:                     8               Gaps:          0

US-09-980-043-4 (1-56) x AF267742 (1-2958)

Qy          1 MetMetGluSerGlyAlaProIleCysHisThrCysGlyGluGlnValGlyHisAspAla 20
              ||||||||||||   ||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGATGGAGTCTAGGTCTCCCATCTGCAACACTTGTGGTGAAGAGATTGGTGCAAAATCA 60

Qy         21 AsnGlyGluLeuPheValAlaCysHisGluCysSerTyrProMetCysLysSerCysPhe 40
              |||||||||   ||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AACGGAGAGTTCTTTGTGGCTTGTCATGAGTGTAGTTTCCCGATCTGCAAAGCTTGTCTT 120

Qy         41 GluPheGluIleLysGluGlyArgLysValCysLeuArgCysGlySer 56
              |||||||||   ||||||||||||||||||||||||||||||||||||
Db        121 GAGTATGAATTCAAAGAAGGTCGAAGAATTTGCTTGCGTTGCGGCAAT 168

RESULT 7
AY099636
LOCUS       AY099636                3109 bp    mRNA    linear   PLN 06-MAY-2002
DEFINITION  Arabidopsis thaliana cellulose synthase-like protein (At4g18780)
            mRNA, complete cds.
ACCESSION   AY099636
VERSION     AY099636.1  GI:20466339
KEYWORDS    FLI_CDNA.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 3109)
  AUTHORS   Nguyen,M., Karlin-Neumann,G., Southwick,A., Lam,B., Miranda,M.,
            Palm,C.J., Bowser,L., Jones,T., Banh,J., Carninci,P., Chen,H.,
            Cheuk,R., Chung,M.K., Hayashizaki,Y., Ishida,J., Kamiya,A.,
            Kawai,J., Kim,C., Lin,J., Liu,S.X., Narusaka,M., Pham,P.K.,
            Sakano,H., Sakurai,T., Satou,M., Seki,M., Shinn,P., Yamada,K.,
            Shinozaki,K., Ecker,J., Theologis,A. and Davis,R.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-APR-2002) DNA Sequencing and Technology Center,
            Stanford University, 855 California Avenue, Palo Alto, CA 94304,
            USA
COMMENT     e mail for correspondence  arabasequence.stanford.edu

            RIKEN Genomic Sciences Center (GSC) members carried out the
            collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN
            Arabidopsis  Full-Length cDNA'). Seki,M., Narusaka,M., Ishida,J.,
            Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J.,
            Hayashizaki,Y. and Shinozaki,K.

            The Salk, Stanford, PGEC (SSP) Consortium members carried out the
            sequencing and annotation of the RAFL cDNAs: Nguyen,M.,
            Southwick,A., Karlin-Neumann,G., Lam,B., Miranda,M., Palm,C.J.,
            Bowser,L., Jones,T., Banh,J., Chen,H., Cheuk,R., Chung,M.K.,
            Kim,C., Lin,J., Liu,S.X., Pham,P.K., Sakano,H., Shinn,P.,
            Yamada,K., Ecker,J., Theologis,A  and Davis,R.W.

            Nguyen,M. (SSP/Stanford) and Seki,M. (RIKEN GSC) contributed
            equally to this work.  Shinozaki,K. (RIKEN GSC) and Davis,R.W.
            (SSP/Stanford)  contributed equally to this work as PIs.
FEATURES             Location/Qualifiers
     source          1. .3109
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
                     /chromosome="4"
                     /clone="RAFL09-65-M12 (R24904)"
                     /note="This clone is in pBluescript
                     ecotype: Columbia"
     gene            1. .3109
                     /gene="At4g18780"
                     /note="synonym: F28A21.190"
     CDS             68. .3025
                     /gene="At4g18780"
                     /codon_start=1
                     /product="cellulose synthase-like protein"
                     /protein_id="AAM20487.1"
                     /db_xref="GI:20466340"
                     /translation="MMESRSPICNTCGEEIGVKSNGEFFVACHECSFPICKACLEYEF
                     KEGRRICLRCGNPYDENVFDDVETKTSKTQSIVPTQTNNTSQDSGIHARHISTVSTID
                     SELNDEYGNPIWKNRVESWKDKKDKKSKKKKKDPKATKAEQHEAQIPTQQHMEDTPPN
                     TESGATDVLSVVIPIPRTKITSYRIVIIMRLIILALFFNYRITHPVDSAYGLWLTSVI
                     CEIWFAVSWVLDQFPKWSPINRETYIDRLSARFEREGEQSQLAAVDFFVSTVDPLKEP
                     PLITANTVLSILALDYPVDKVSCYVSDDGAAMLSFESLVETADFARKWVPFCKKYSIE
                     PRAPEFYFSLKIDYLRDKVQPSFVKERRAMKRDYEEFKIRMNALVAKAQKTPEEGWTM
                     QDGTSWPGNNTRDHPGMIQVFLGYSGARDIEGNELPRLVYVSREKRPGYQHHKKAGAE
                     NALVRVSAVLTNAPFILNLDCDHYVNNSKAVREAMCFLMDPVVGQDVCFVQFPQRFDG
                     IDKSDRYANRNIVFFDVNMRGLDGIQGPVYVGTGTVFRRQALYGYSPPSKPRILPQSS
                     SSSCCCLTKKKQPQDPSEIYKDAKREELDAAIFNLGDLDNYDEYDRSMLISQTSFEKT
                     FGLSTVFIESTLMENGGVPDSVNPSTLIKEAIHVISCGYEEKTEWGKEIGWIYGSITE
                     DILTGFKMHCRGWRSIYCMPLRPAFKGSAPINLSDRLHQVLRWALGSVEIFLSRHCPL
                     WYGCSGGRLKLLQRLAYINTIVYPFTSLPLVAYCTLPAICLLTGKFIIPTLSNLASML
                     FLGLFISIILTSVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQGFLKMLAGLDT
                     NFTVTSKTADDLEFGELYIVKWTTLLIPPTSLLIINLVGVVAGFSDALNKGYEAWGPL
                     FGKVFFAFWVILHLYPFLKGLMGRQNRTPTIVILWSILLASVFSLVWVRINPFVSKTD
                     TTSLSLNCLLIDC"
BASE COUNT      859 a    631 c    687 g    932 t
ORIGIN

Alignment Scores:
Pred. No.:              6.98e-20        Length:        3109
Score:                  246.00          Matches:       38
Percent Similarity:     89.29%          Conservative:  12
Best Local Similarity:  67.86%          Mismatches:    6
Query Match:            75.93%          Indels:        0
DB:                     8               Gaps:          0

US-09-980-043-4 (1-56) x AY099636 (1-3109)

Qy          1 MetMetGluSerGlyAlaProIleCysHisThrCysGlyGluGlnValGlyHisAspAla 20
              ||||||||||||   ||||||||||:||||||||||||:::::|||          :::
Db         68 ATGATGGAGTCTAGGTCTCCCATCTGCAACACTTGTGGTGAAGAGATTGGTGTAAAATCA 127

Qy         21 AsnGlyGluLeuPheValAlaCysHisGluCysSerTyrProMetCysLysSerCysPhe 40
              |||||||||   ||||||||||||||||||||||||:::|||:::||||||:::|||
Db        128 AACGGAGAGTTCTTTGTGGCTTGTCATGAGTGTAGTTTCCCAATCTGCAAAGCTTGTCTT 187

Qy         41 GluPheGluIleLysGluGlyArgLysValCysLeuArgCysGlySer 56
              |||:::|||   |||||:|||||:::::||||||||||||||:::
Db        188 GAGTATGAATTCAAAGAAGGTCGAAGAATTTGCTTGCGTTGCGGCAAT 235

RESULT 8
ATF28A21/c
LOCUS       ATF28A21                94301 bp    DNA     linear   PLN 08-SEP-1999
DEFINITION  Arabidopsis thaliana DNA chromosome 4, BAC clone F28A21 (ESSA
            project).
ACCESSION   AL035526
VERSION     AL035526.1  GI:4539378
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Bevan,M., Mueller,M.W., Muendlein,A., Felber,R., Bancroft,I.,
            Mewes,H.W., Mayer,K.F.X., Lemcke,K. and Schueller,C.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 94301)
  AUTHORS   EU Arabidopsis sequencing,project.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-SEP-1999) MIPS, at the Max-Planck-Institut fuer
            Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
            schuelle@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de Project
            Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
```

```
            Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
            E-mail: michael.bevan@bbsrc.ac.uk
COMMENT     Information on performance of analysis and a more detailed
            annotation of this entry and other sequences of chromosomes 3, 4
            and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/.
EATURES             Location/Qualifiers
     source         1. .94301
                    /organism="Arabidopsis thaliana"
                    /mol_type="genomic DNA"
                    /variety="Columbia"
                    /db_xref="taxon:3702"
                    /chromosome="4"
     gene           9652. .10227
                    /gene="F28A21.20"
     CDS            9652. .10227
                    /gene="F28A21.20"
                    /note="similarity to predicted proteins, Arabidopsis
                    thaliana
                    contains EST gb:T46643, R29928"
                    /codon_start=1
                    /product="putative protein"
                    /protein_id="CAB37446.1"
                    /db_xref="GI:4539380"
                    /translation="MSSDRHTPTKDPPDHPSSSSNHHKQPLPPQPQQPLSRYESQKRR
                    DWNTFVQYLKSQNPPLMMSQFDYTHVLSFLRYLDQFGKTKVHHQACVFFGQPDPPGPC
                    TCPLKQAWGSLDALIGRLRAAYEEHGGGSPDTNPFANGSIRVHLREVRESQAKARGIP
                    YRKKKRRKTKNEVVVVKKDVANSSTPNQSFT"
     exon           9652. .10227
                    /gene="F28A21.20"
                    /number=1
     gene           13217. .13711
                    /gene="F28A21.30"
     CDS            13217. .13711
                    /gene="F28A21.30"
                    /note="similarity to various predicted proteins,
                    Arabidopsis thaliana"
                    /codon_start=1
                    /product="putative protein"
                    /protein_id="CAB37447.1"
                    /db_xref="GI:4539381"
                    /translation="MESSKQKRCRSSVVETIEAPLPLVWSILRSFDKPQAYQRFVKSC
                    TMRSGGGGGKGGEGKGSVRDVTLVSGFPADFSTERLEELDDESHVMVVSIIGGNHRLV
                    NYKSKTKVVASPEDMAKKTVVVESYVVDVPEGTSEEDTIFFVDNIIRYNLTSLAKLTK
                    KMMK"
     exon           13217. .13711
                    /gene="F28A21.30"
                    /number=1
     gene           16036. .17454
                    /gene="F28A21.40"
     CDS            16036. .17454
                    /gene="F28A21.40"
                    /note="similarity to predicted protein, Arabidopsis
                    thaliana"
                    /codon_start=1
                    /product="putative protein"
                    /protein_id="CAB37448.1"
                    /db_xref="GI:4539382"
                    /translation="MEGKKLNLYAPLPSIRRIPSMIERSSELENKKTTITRPELRSCP
                    STKQETPVFVLPDQSFDHLTEPASIPFMWEQIPGKPKDDMATLIQESGLLETDDEEED
                    EVDEDLDTVSSNVSPSVNCSTSGVSEIEKTGERSDCDEKSRESLDLMMSRFLPAAKAM
                    ALQTHQKHQSSYNSSEQKLITQNREALVARQRSQLVAEHEHFAIVQSLYDDLNIDDDT
                    EDDENDDDGDDYGHVDHKIYPAEVTKKACGFLPRLCAKNSFKFLNPVPLDSVSNRSMK
                    HSGDQTSLPNWSTRRLSGFISPYRTSCSDSGFLGTPEKPESFKRLNRGISKSQELYPT
                    RTRREVLPNYSNSGEIKMFRNSISTPSRIQRTTIHDSRFLVEEVNRRRNNKSGNLQQT
                    SSEYTPALSPPPLPETPSRSWLGRTLLPPVNPRPYGVVLGQVGIKKLNQEVLESTKWE
                    TIVKTSYVHNDHARYSQVLITL"
     exon           16036. .17454
                    /gene="F28A21.40"
                    /number=1
     gene           19501. .22435
                    /gene="F28A21.50"
     CDS            join(<19501. .19652,19742. .19813,19902. .20045,20144. .20287,
                    20375. .20469,20559. .20987,21071. .21261,21342. .21474,
                    21531. .21709,21893. .22084,22178. .22435)
                    /gene="F28A21.50"
                    /codon_start=1
                    /product="putative protein (fragment)"
                    /protein_id="CAB37449.1"
                    /db_xref="GI:4539383"
                    /translation="FVLNSGSALLKFRARVNSDPHGTLANWNVSGINDLCYWSGVTCV
                    DGKVQILDLSGYSLEGTLAPELSQLSDLRSLILSRNHFSGGIPKEYGSFENLEVLDLR
                    ENDLSGQIPPELSNGLSLKHLLLSGNKFSDDMRIKIVRLQSSYEVRLKKSPKLSPLAV
                    LGCINRKLGHCVSRNFIIQVKKVEAIVFRIKATSRRFLKAFPSFLEETDIYKRRELLE
                    ETSNLAAEPAPSAPSPSPGIITEASPRSSGSFPAVTNAKKRRPPLVPPVPSPDKGSTS
                    PDISKNQPQDNKQSKGSKHVWLYVVIAVASFVGLLIIVAVIFFCRKRAVKSIGPWKTG
                    LSGQLQKAFVTGVPKLNRSELETACEDFSNIIETFDGYTVYKGTLSSGVEIAVASTAI
                    AESKEWTRAMEMAYRRKIDTLSRINHKNFVNLIGYCEEDDPFNRMMVFEYAPNGTLFE
                    HLHVNNVFRLSDKETEHLDWSARMRIIMGTAYCLQHMHGMNPPMAHTDFNSSEIYLTD
                    DYAAKVSEIPFNLEARLNPKKHVSGDLEQTSLLLPPEPEANVHSFGVLMLEIISGKLS
                    FSDEYGSIEQWASKYLEKDDLGEMIDPSLKTFKEEELEVICDVIRECLKTEQRQRPSM
                    KDVAEQLKQVINITPEKATPRSSPLWWAELEILSSEAT"
     exon           19501. .19652
                    /gene="F28A21.50"
                    /number=1
     intron         19653. .19741
                    /gene="F28A21.50"
                    /number=1
     exon           19742. .19813
                    /gene="F28A21.50"
                    /number=2
     intron         19814. .19901
                    /gene="F28A21.50"
                    /number=2
     exon           19902. .20045
                    /gene="F28A21.50"
                    /number=3
     intron         20046. .20143
                    /gene="F28A21.50"
                    /number=3
     exon           20144. .20287
                    /gene="F28A21.50"
                    /number=4
     intron         20288. .20374
                    /gene="F28A21.50"
                    /number=4
     exon           20375. .20469
                    /gene="F28A21.50"
                    /number=5
     intron         20470. .20558
                    /gene="F28A21.50"
                    /number=5
     exon           20559. .20987
                    /gene="F28A21.50"
                    /number=6
     intron         20988. .21070
                    /gene="F28A21.50"
                    /number=6
     exon           21071. .21261
                    /gene="F28A21.50"
                    /number=7
     intron         21262. .21341
                    /gene="F28A21.50"
                    /number=7
     exon           21342. .21474
                    /gene="F28A21.50"
                    /number=8
     intron         21475. .21530
                    /gene="F28A21.50"
                    /number=8
     exon           21531. .21709
                    /gene="F28A21.50"
                    /number=9
     intron         21710. .21892
                    /gene="F28A21.50"
                    /number=9
     exon           21893. .22084
```

```
                     /gene="F28A21.50"
                     /number=10
     intron          22085. .22177
                     /gene="F28A21.50"
                     /number=10
     exon            22178. .22435
                     /gene="F28A21.50"
                     /number=11
     gene            22901. .24271
                     /gene="F28A21.60"
     gene            complement(join(22901. .23176,23858. .24271))
                     /gene="F28A21.60"
     CDS             complement(join(22901. .23176,23858. .24271))
                     /gene="F28A21.60"
                     /note="similarity to transcription factor OBF3.1, Zea
                     mays, PIR2:S33223"
                     /codon_start=1
                     /product="putative protein"
                     /protein_id="CAB37450.1"
                     /db_xref="GI:4539384"
                     /translation="MRNLVEEKFLEFYESWVIQLELYLHQLLIAHNNNTMSETELRHL
                     ISKLTTHHKAYYTAKWAAIREDVLAFFGSVWLNPLENACSWLTGWKPSMVFRMVDRLR
                     KSRVVLVEAQVKKLEELRVKTKFDEQKIEREMERYQVAMADRKMVELARLGCHVGGES
                     VMVVEAAVRGLSMGLEKMVKAADCVRLKTLKGILDILTPPQCVEFLAAAATFQVQLRR
                     WGNRRHYVTHS"
     exon            complement(22901. .23176)
                     /gene="F28A21.60"
                     /number=1
     intron          complement(23177. .23857)
                     /gene="F28A21.60"
                     /number=1
     exon            complement(23858. .24271)

Alignment Scores:
Pred. No.:              2.6e-18         Length:         94301
Score:                  246.00          Matches:        38
Percent Similarity:     89.29%          Conservative:   12
Best Local Similarity:  67.86%          Mismatches:     6
Query Match:            75.93%          Indels:         0
DB:                     8               Gaps:           0

Us-09-980-043-4 (1-56) x ATF28A21 (1-94301)

Qy          1 MetMetGluSerGlyAlaProIleCysHisThrCysGlyGluGlnValGlyHisAspAla 20
              |||||||||||||   :::|||||: ||::: |||||:|||' :::::|||     :::
Db      75577 ATGATGGAGTCTAGGTCTCCCATTTGCAACACTTGTGGTGAAGAGATTGGTGTAAAATCA 75518

Qy         21 AsnGlyGluLeuPheValAlaCysHisGluCysSerTyrProMetCysLysSerCysPhe 40
              |||||||||   |||||||||||||||||||||:::|' :::|:||||:::|||
Db      75517 AACGGAGAGTTCTTTGTGGCTTGTCATGAGTGTAGTTTCCCGATCTGCAAAGCTTGTCTT 75458

Qy         41 GluPheGluIleLysGluGlyArgLysValCysLeuArgCysGlySer 56
              |||:::|||   |||||||||||:::::|||||||||||' ||:::
Db      75457 GAGTATGAATTCAAAGAAGGTCGAAGAATTTGCTTGCGTTGCGGCAAT 75410

RESULT 9
ATCHRIV49/c
LOCUS       ATCHRIV49              199075 bp    DNA     linear   PLN 16-MAR-2000
DEFINITION  Arabidopsis thaliana DNA chromosome 4, contig fragment No. 49.
ACCESSION   AL161549
VERSION     AL161549.2  GI:7268649
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 28924)
  AUTHORS   Hilbert,H., Braun,M., Holzer,E., Brandt,A., Duesterhoeft,A.,
            Mewes,H.W., Lemcke,K. and Mayer,K.F.X.
  JOURNAL   Unpublished
REFERENCE   2  (bases 19282 to 113583)
  AUTHORS   Mueller,M.W., Muendlein,A., Felber,R., Mewes,H.W., Lemcke,K. and
            Mayer,K.F.X.
  JOURNAL   Unpublished
REFERENCE   3  (bases 113371 to 118148)
  AUTHORS   Robben,J., Grymonprez,B., Volckaert,G., Mewes,H.W., Lemcke,K. and
            Mayer,K.F.X.
  JOURNAL   Unpublished
REFERENCE   4  (bases 117865 to 199075)
  AUTHORS   Pohl,T., Weizenegger,T., Mewes,H.W., Lemcke,K. and Mayer,K.F.X.
  JOURNAL   Unpublished
REFERENCE   5  (bases 1 to 199075)
  AUTHORS   EU Arabidopsis sequencing,project.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer
            Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
            lemcke@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de Project
            Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
            Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
            E-mail: michael.bevan@bbsrc.ac.uk
COMMENT     Information on performance of analysis and a more detailed
            annotation of this entry and other sequences of chromosomes 3, 4
            and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/
            this fragment has an overlap with ATCHRIV48 at the 5' end and an
            overlap with ATCHRIV50 at the 3' end.
FEATURES             Location/Qualifiers
     source          1. .199075
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /variety="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="4"
     gene            9577. .12672
                     /gene="AT4g18560"
     CDS             join(9577. .9906,9958. .10809,10980. .11111,11160. .11377,
                     11463. .11574,12310. .12447,12541. .12672)
                     /gene="AT4g18560"
                     /note="similarity to pherophorin-S, Volvox carteri,
                     PIR:T10798
                     contains EST gb:AI997040.1, W43771"
                     /codon_start=1
                     /product="pherophorin-like protein"
                     /protein_id="CAB78858.1"
                     /db_xref="GI:7268650"
                     /translation="MVAGKVRVTMGFHKSPSTKKTKDMPSPLPLPPPPPPPLKPPSSG
                     SATTKPPINPSKPGFTRSFGVYFPRASAQVHATAAAASHNGVVSELRRQVEELREREA
                     LLKTENLEIADKNGEIDELRKETARLAEDNERLRREFDRSEEMRRECETREKEMEAEI
                     VELRKLVSSESDDHALSVSQRFQGLMDVSAKSNLIRSLKRVGSLRNLPEPITNQENTN
                     KSISSSGDADGDIYRKDEIESYSRSSNSEELTESSSLSTVRSRVPRVPKPPPKRSISL
                     GDSTENRADPPPQKSIPPPPPPPPPLLQQPPPPPSVSKAPPPPPPPPPPKSLSIASA
                     KVRRVPEVVEFYHSLMRRDSTNSRRDSTGGGNAAAEAILANSNARDMIGEIENRSVYL
                     LAIKTDVETQGDFIRFLIKEVGNAAFSDIEDVVPPVKWLDDELSYLFKVCKFVVVSLK
                     VDERAVLKHFEWPEQKADALREAAFCYFDLKKLISEASRFREDPRQSSSSALKKMQAL
                     FEKLEHGVYSLSRMRESAATKFKSFQIPVDWMLETGITSQIKLASVKLAMKYMKRVSA
                     ELEAIEGGGPEEEELIVQGVRFAFRVHQFAGGFDAETMKAFEELRDKARSCHVQCQSQ
                     THQHKLCFRSTPC"
     exon            9577. .9906
                     /gene="AT4g18560"
                     /number=1
     gene            9612. .11059
                     /gene="AT4g18570"
     gene            complement(join(9612. .9715,10483. .10747,10973. .11059))
                     /gene="AT4g18570"
     CDS             complement(join(9612. .9715,10483. .10747,10973. .11059))
                     /gene="AT4g18570"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAB78859.1"
                     /db_xref="GI:7268651"
                     /translation="MSENAAFPTSLIKNLMKSPCVSTSVFICNQVLLPRRRFHRRWNL
                     FSSLWSLFASTSDKTQPLPVLFSLSPTQCLNSSAAVVEEAEEELWKQTEVVVVVEEAV
                     AEVEEEEAESISAVAVAEPEEGGLSGGGGGGGRGKGEGMSLVFLVEGDL"
     exon            complement(9612. .9715)
                     /gene="AT4g18570"
                     /number=1
```

```
intron          complement(9716. .10482)
                /gene="AT4g18570"
                /number=1
intron          9907. .9957
                /gene="AT4g18570"
                /number=1
exon            9958. .10809
                /gene="AT4g18570"
                /number=2
exon            complement(10483. .10747)
                /gene="AT4g18570"
                /number=2
intron          complement(10748. .10972)
                /gene="AT4g18570"
                /number=2
intron          10810. .10979
                /gene="AT4g18570"
                /number=2
exon            complement(10973. .11059)
                /gene="AT4g18570"
                /number=3
exon            10980. .11111
                /gene="AT4g18560"
                /number=3
intron          11112. .11159
                /gene="AT4g18560"
                /number=3
exon            11160. .11377
                /gene="AT4g18560"
                /number=4
intron          11378. .11462
                /gene="AT4g18560"
                /number=4
exon            11463. .11574
                /gene="AT4g18560"
                /number=5
intron          11575. .12309
                /gene="AT4g18560"
                /number=5
exon            12310. .12447
                /gene="AT4g18560"
                /number=6
intron          12448. .12540
                /gene="AT4g18560"
                /number=6
exon            12541. .12672
                /gene="AT4g18560"
                /number=7
gene            12991. .13662
                /gene="AT4g18580"
gene            complement(join(12991. .13077,13176. .13264,13382. .13507,
                13596. .13662))
                /gene="AT4g18580"
CDS             complement(join(12991. .13077,13176. .13264,13382. .13507,
                13596. .13662))
                /gene="AT4g18580"
                /note="contains EST gb:AI998801.1, AI138148"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="CAB78860.1"
                /db_xref="GI:7268652"
                /translation="MATLAPVHAVSVLHSRKRGGCNGNFVSFSGVPLPYLIRPRRKIP
                PVDTFDPYNDSDDESGEGDKGHSDDEYVVVDKAEAIMPNVKTEKPGAIQMMELSKDAV
                GDDESKAKDAGYSSDEWFVV"
exon            complement(12991. .13077)
                /gene="AT4g18580"
                /number=1
intron          complement(13078. .13175)
                /gene="AT4g18580"
                /number=1
exon            complement(13176. .13264)
                /gene="AT4g18580"
                /number=2
intron          complement(13265. .13381)
                /gene="AT4g18580"
                /number=2
exon            complement(13382. .13507)
                /gene="AT4g18580"
                /number=3
intron          complement(13508. .13595)
                /gene="AT4g18580"
                /number=3
exon            complement(13596. .13662)
                /gene="AT4g18580"
                /number=4
gene            14462. .20681
                /gene="AT4g18590"
CDS             join(14462. .14817,15014. .15052,16278. .16535,16652. .16752,
                17087. .17423,17529. .17854,18312. .18366,19447. .19527,
                19615. .20681)
                /gene="AT4g18590"
                /note="overlap to next genmodel
                similarity to 1% of protein precursor, pollen, Zea mays,
                PIR2:S28108
                contains pollen proline-rich 2 family signature
                AA309 5.8
                contains EST gb:AI899674 ... AA721906"
                /codon_start=1
                /product="pollen-specific protein-like"
                /protein_id="CAB78861.1"
                /db_xref="GI:7268653"
                /translation="MDTSSPAAFVNGALLRRYIGQKVRAVIQVIRSDVGSVIGKSTDD
                QQIVVKGSPQPPLTTYLEVIGIAETRNTIRAEVWTNFGDSFDVQNYNELCKLANDRRK
                AMTDLQMEVEVDTNSSLQESLPKPQVMYRCKKCRRIVAIEENIVPHEPGKGEECFAWK
                KRSGNSEQVQCSSIFVEPMKWMQTIEDGMVEEKLLCFGCNGRLGYFNWAGMQCSCGAW
                LKITDGFNFMILYKTELCGSCELVVLPSNISNKENDYKREEVSKRRPESFNILKQMAS
                KAIFFFFVSAVCLSSLAGVAIADADDFDRFQIQGSVYCDTCRVQFVTRLSKFLEGAKV
                KLECRSRTNGTVTLTKEAVTDKTGSYRMEVTGDHEEEVCELVLVESPDSGCSDVSKEA
                YLRNAAKISLTANDGIVSHETRIVNPLGFMVQTPSAECPAAFKELGIVPDGLLFGWFG
                GQARMISETDGRSVTTHGAATCISGSFKEGYVESGTSAKARYAAAGSFGRIFSLGTGS
                SGKHLLCEPVSVVKLGTPEPVIFSGKTASVEDRSVVIVSLGKVLLCGLVSVVELEASE
```

```
Alignment Scores:
Pred. No.:              5.74e-18        Length:         199075
Score:                  246.00          Matches:        38
Percent Similarity:     89.29%          Conservative:   12
Best Local Similarity:  67.86%          Mismatches:     6
Query Match:            75.93%          Indels:         0
DB:                     8               Gaps:           0

US-09-980-043-4 (1-56) x ATCHRIV49 (1-199075)

Qy          1 MetMetGluSerGlyAlaProIleCysHisThrCysGlyGluGlnValGlyHisAspAla 20
              |||||||||||||   :::|||||||||:::|||:|||||||||:::::|||      :::
Db      94858 ATGATGGAGTCTAGGTCTCCCATCTGCAACACTTGTGGTGAAGAGATTGGTGTAAAATCA 94799

Qy         21 AsnGlyGluLeuPheValAlaCysHisGluCysSerTyrProMetCysLysSerCysPhe 40
              |||||||||   |||||||||||||||||:||||:::|||:::|||||:::|||
Db      94798 AACGGAGAGTTCTTTGTGGCTTGTCATGAGTGTAGTTTCCCGATCTGCAAAGCTTGTCTT 94739

Qy         41 GluPheGluIleLysGluGlyArgLysValCysLeuArgCysGlySer 56
              |||:::|||   ||||||||||,|::::::|||:|||||||||||:::
Db      94738 GAGTATGAATTCAAAGAAGGTCGAAGAATTTGCTTGCGTTGCGGCAAT 94691

RESULT 10
AF139442
LOCUS       AF139442                1095 bp    DNA     linear   PLN 12-JUL-2002
DEFINITION  Gossypium herbaceum cellulose synthase A1 (CelA1) gene, partial
            cds.
ACCESSION   AF139442
VERSION     AF139442.1  GI:4928981
KEYWORDS    .
SOURCE      Gossypium herbaceum
  ORGANISM  Gossypium herbaceum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
```

```
          rosids; eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium.
REFERENCE   1  (bases 1 to 1095)
  AUTHORS   Cronn,R.C., Small,R.L. and Wendel,J.F.
  TITLE     Duplicated genes evolve independently after polyploid formation in
            cotton
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (25), 14406-14411 (1999)
  MEDLINE   20056257
   PUBMED   10588718
REFERENCE   2  (bases 1 to 1095)
  AUTHORS   Cronn,R.C. and Wendel,J.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-MAR-1999) Botany, Iowa State University, 353 Bessey,
            Ames, IA 50011, USA
FEATURES             Location/Qualifiers
     source          1. .1095
                     /organism="Gossypium herbaceum"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:34274"
                     /map="Homoeologous Assemblage 10"
                     /note="maps to homoeologous assemblage 10; refer to
                     Brubaker et al. (Genome, 1999) for details."
     gene            <1. .>1095
                     /gene="CelA1"
     mRNA            join(<1. .151,324. .397,484. .535,628. .772,903. .>1095)
                     /gene="CelA1"
                     /product="cellulose synthase A1"
     CDS             join(<1. .151,324. .397,484. .535,628. .772,903. .>1095)
                     /gene="CelA1"
                     /codon_start=1
                     /product="cellulose synthase A1"
                     /protein_id="AAD33795.1"
                     /db_xref="GI:4928982"
                     /translation="HTCGEHVGLNVNGEPFVACHECNFPICKSCFEYDLKEGRKACLR
                     CGSPYDENLLDDVEKATGDQSTMAAHLSKSQDVGIHARHISSVSTLDSEMTEDNGNPI
                     WKNRVESWKEKKNKKKKPATTKVEREAEIPPEQQMEDKPAPDASQPLSTIIPIPKSRL
                     APYRTVIIMRLIILGLFFHYRVTNPVDSAFGLWLTSVICEIWFAF"
BASE COUNT      315 a    180 c    238 g    362 t
ORIGIN

Alignment Scores:
Pred. No.:                3.15e-17        Length:          1095
Score:                    220.00          Matches:         35
Percent Similarity:       89.36%          Conservative:    7
Best Local Similarity:    74.47%          Mismatches:      5
Query Match:              67.90%          Indels:          0
DB:                       8               Gaps:            0

US-09-980-043-4 (1-56) x AF139442 (1-1095)

Qy          10 HisThrCysGlyGluGlnValGlyHisAspAlaAsnGlyGluLeuPheValAlaCysHis 29
               ||||||||||||||||   |||.||   :::   |||  :||   ||||||||||||||||
Db           1 CACACTTGTGGTGAACATGTTGGGTTGAATGTTAATGGTGAACCCTTTGTGGCTTGCCAT 60

Qy          30 GluCysSerTyrProMetCysLysSerCysPheGluPheGluIleLysGluGlyArgLys 49
               ||||||::::::|||:::|||||||||||||'||||||||||:||:||||||||||||
Db          61 GAATGTAATTTCCCTATTTGTAAGAGTTGTTTTGAGTATGATGTTAAGGAAGGACGAAAA 120

Qy          50 ValCysLeuArgCysGlySer 56
                   |||||||||||||||||
Db         121 GCTTGCTTGCGTTGTGGTAGT 141

RESULT 11
AF139443
LOCUS       AF139443                1095 bp    DNA     linear   PLN 12-JUL-2002
DEFINITION  Gossypium hirsutum cellulose synthase A1 (CelA1) gene, partial cds.
ACCESSION   AF139443
VERSION     AF139443.1  GI:4928983
KEYWORDS    .
SOURCE      Gossypium hirsutum (upland cotton)
  ORGANISM  Gossypium hirsutum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium.
REFERENCE   1  (bases 1 to 1095)
  AUTHORS   Cronn,R.C., Small,R.L. and Wendel,J.F.
  TITLE     Duplicated genes evolve independently after polyploid formation in
            cotton
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (25), 14406-14411 (1999)
  MEDLINE   20056257
   PUBMED   10588718
REFERENCE   2  (bases 1 to 1095)
  AUTHORS   Cronn,R.C. and Wendel,J.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-MAR-1999) Botany, Iowa State University, 353 Bessey,
            Ames, IA 50011, USA
FEATURES             Location/Qualifiers
     source          1. .1095
                     /organism="Gossypium hirsutum"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3635"
                     /map="Homoeologous Assemblage 10"
                     /note="maps to homoeologous assemblage 10; refer to
                     Brubaker et al. (Genome, 1999) for details."
     gene            <1. .>1095
                     /gene="CelA1"
     mRNA            join(<1. .151,324. .397,484. .535,628. .772,903. .>1095)
                     /gene="CelA1"
                     /product="cellulose synthase A1"
     CDS             join(<1. .151,324. .397,484. .535,628. .772,903. .>1095)
                     /gene="CelA1"
                     /codon_start=1
                     /product="cellulose synthase A1"
                     /protein_id="AAD33796.1"
                     /db_xref="GI:4928984"
                     /translation="HTCGEHVGLNVNGEPFVACHECNFPICKSCFEYDLKEGRKACLR
                     CGSPYDENLLDDVEKATGDQSTMAAHLSKSQDVGIHARHISSVSTLDSEMTEDNGNPI
                     WKNRVESWKEKKNKKKKPATTKVEREAEIPPEQQMEDKPAPDASQPLSTIIPIPKSRL
                     APYRTVIIMRLIILGLFFHYRVTNPVDSAFGLWLTSVICEIWFAF"
BASE COUNT      313 a    179 c    240 g    363 t
ORIGIN

Alignment Scores:
Pred. No.:                3.15e-17        Length:          1095
Score:                    220.00          Matches:         35
Percent Similarity:       89.36%          Conservative:    7
Best Local Similarity:    74.47%          Mismatches:      5
Query Match:              67.90%          Indels:          0
DB:                       8               Gaps:            0

US-09-980-043-4 (1-56) x AF139443 (1-1095)

Qy          10 HisThrCysGlyGluGlnValGlyHisAspAlaAsnGlyGluLeuPheValAlaCysHis 29
               ||||||||||||||||   ||||||   :::   |||||||||   ||||||||||||||||
Db           1 CACACTTGTGGTGAACATGTTGGGTTGAATGTTAATGGTGAACCCTTTGTGGCTTGCCAT 60

Qy          30 GluCysSerTyrProMetCysLysSerCysPheGluPheGluIleLysGluGlyArgLys 49
               ||||||::::::|||:::|||||||||||||||||||:::::::::|||||||||||||
Db          61 GAATGTAATTTCCCTATTTGTAAGAGTTGTTTTGAGTATGATCTTAAGGAAGGACGAAAA 120

Qy          50 ValCysLeuArgCysGlySer 56
                   |||||||||||||||||
Db         121 GCTTGCTTGCGTTGTGGTAGT 141

RESULT 12
AF139446
LOCUS       AF139446                1096 bp    DNA     linear   PLN 01-JUN-1999
DEFINITION  Gossypium robinsonii cellulose synthase A1 (CelA1) gene, partial
            cds.
ACCESSION   AF139446
VERSION     AF139446.1  GI:4928989
KEYWORDS    .
SOURCE      Gossypium robinsonii
  ORGANISM  Gossypium robinsonii
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium.
```

```
EFERENCE   1  (bases 1 to 1096)
 AUTHORS   Cronn,R.C., Small,R.L. and Wendel,J.F.
 TITLE     Duplicated genes evolve independently following polyploid formation
           in cotton
 JOURNAL   Unpublished
EFERENCE   2  (bases 1 to 1096)
 AUTHORS   Cronn,R.C. and Wendel,J.F.
 TITLE     Direct Submission
 JOURNAL   Submitted (30-MAR-1999) Botany, Iowa State University, 353 Bessey,
           Ames, IA 50011, USA
EATURES             Location/Qualifiers
     source         1. .1096
                    /organism="Gossypium robinsonii"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:3636"
                    /map="Homoeologous Assemblage 10"
     gene           <1. .>1096
                    /gene="CelA1"
     mRNA           join(<1. .151,325. .398,485. .536,630. .774,904. .>1096)
                    /gene="CelA1"
                    /product="cellulose synthase A1"
     CDS            join(<1. .151,325. .398,485. .536,630. .774,904. .>1096)
                    /gene="CelA1"
                    /codon_start=1
                    /product="cellulose synthase A1"
                    /protein_id="AAD33799.1"
                    /db_xref="GI:4928990"
                    /translation="HTCGEHVGLNVNGEPFVACHECNFPICKSCFEYDLKEGRKACLR
                    CGSPYDENLLDDVEKATGDQSTMAAHLSKSQDIGIHARHISSVSTLDSEMTEDNGNPI
                    WKNRVESWKEKKNKKKQPATTKVEREAEIPPEQQMEDKPAPDASQPLSTIIPIPKSRL
                    APYRTVIIMRLIILGLFFHYRVTNPVDSAFGLWLTSVICEIWFAF"
ASE COUNT      315 a    184 c    237 g    359 t      1 others
RIGIN

lignment Scores:
red. No.:              3.16e-17      Length:        1096
core:                  220.00        Matches:       35
ercent Similarity:     89.36%        Conservative:  7
est Local Similarity:  74.47%        Mismatches:    5
uery Match:            67.90%        Indels:        0
B:                     8             Gaps:          0

S-09-980-043-4 (1-56) x AF139446 (1-1096)

Qy         10 HisThrCysGlyGluGlnValGlyHisAspAlaAsnGlyGluLeuPheValAlaCysHis 29
              ||||||||||||||||   ||||||   :::   |||||||||   ||||||||||||||||
Db          1 CACACTTGTGGTGAACATGTTGGGTTGAATGTTAATGGTGAACCCTTTGTGGCTTGCCAT 60

Qy         30 GluCysSerTyrProMetCysLysSerCysPheGluPheGluIleLysGluGlyArgLys 49
              ||||||:::::|||:::||||||||||||||||||:::::::::|||||||||||||||||
Db         61 GAATGTAATTTCCCTATTTGTAAGAGTTGTTTTGAGTATGATCTTAAGGAAGGACGAAAA 120

Qy         50 ValCysLeuArgCysGlySer 56
              ||||||||||||||||||
Db        121 GCTTGCTTGCGTTGTGGTAGT 141

ESULT 13
F139444
OCUS       AF139444               1097 bp    DNA     linear   PLN 12-JUL-2002
EFINITION  Gossypium raimondii cellulose synthase A1 (CelA1) gene, partial
           cds.
ACCESSION  AF139444
ERSION     AF139444.1  GI:4928985
EYWORDS    .
OURCE      Gossypium raimondii
  ORGANISM Gossypium raimondii
           Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
           Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
           rosids; eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium.
EFERENCE   1  (bases 1 to 1097)
  AUTHORS  Cronn,R.C., Small,R.L. and Wendel,J.F.
  TITLE    Duplicated genes evolve independently after polyploid formation in
           cotton
  JOURNAL  Proc. Natl. Acad. Sci. U.S.A. 96 (25), 14406-14411 (1999)
  MEDLINE  20056257
   PUBMED  10588718
REFERENCE  2  (bases 1 to 1097)
  AUTHORS  Cronn,R.C. and Wendel,J.F.
  TITLE    Direct Submission
  JOURNAL  Submitted (30-MAR-1999) Botany, Iowa State University, 353 Bessey,
           Ames, IA 50011, USA
FEATURES             Location/Qualifiers
     source         1. .1097
                    /organism="Gossypium raimondii"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:29730"
                    /map="Homoeologous Assemblage 10"
                    /note="maps to homoeologous assemblage 10; refer to
                    Brubaker et al. (Genome, 1999) for details."
     gene           <1. .>1097
                    /gene="CelA1"
     mRNA           join(<1. .151,325. .398,485. .536,630. .774,905. .>1097)
                    /gene="CelA1"
                    /product="cellulose synthase A1"
     CDS            join(<1. .151,325. .398,485. .536,630. .774,905. .>1097)
                    /gene="CelA1"
                    /codon_start=1
                    /product="cellulose synthase A1"
                    /protein_id="AAD33797.1"
                    /db_xref="GI:4928986"
                    /translation="HTCGEHVGLNVNGEPFVACHECNFPICKSCFEYDLKEGRKACLR
                    CGSPYDENLLDDVEKATGDQSTMAAHLNKSQDVGIHARHISSVSTLDSEMAEDNGNSI
                    WKNRVESWKEKKNKKKKPATTKVEREAEIPPEQQMEDKPAPDASQPLSTIIPIPKSPI
                    APYRTVIIMRLIILGLFFHYRVTNPVDSAFGLWLTSVICEIWFAF"
BASE COUNT      318 a    181 c    239 g    359 t
ORIGIN

Alignment Scores:
Pred. No.:             3.16e-17      Length:        1097
Score:                 220.00        Matches:       35
Percent Similarity:    89.36%        Conservative:  7
Best Local Similarity: 74.47%        Mismatches:    5
Query Match:           67.90%        Indels:        0
DB:                    8             Gaps:          0

US-09-980-043-4 (1-56) x AF139444 (1-1097)

Qy         10 HisThrCysGlyGluGlnValGlyHisAspAlaAsnGlyGluLeuPheValAlaCysHis 29
              ||| |||||||:|   ||||||   :::   ||.|||||   |||||||||||||||||
Db          1 CACACTTGTGGTGAACATGTTGGGTTGAATGTTAATGGTGAACCTTTTGTGGCTTGCCAT 60

Qy         30 GluCysSerTyrProMetCysLysSerCysPheGluPheGluIleLysGluGlyArgLys 49
              |||:||:::::::|:::|||||||||||,|||||:::::::::||||||||||||||||
Db         61 GAATGTAATTTCCCTATTTGTAAGAGCTGTTTTGAGTATGATCTTAAGGAAGGACGAAAA 120

Qy         50 ValCysLeuArgCysGlySer 56
              |||||||||| ||||||
Db        121 GCTTGCTTGCGTTGTGGTAGT 141

RESULT 14
AF139445
LOCUS       AF139445               1097 bp    DNA     linear   PLN 12-JUL-2002
DEFINITION  Gossypium hirsutum D subgenome cellulose synthase A1 (CelA1) gene,
            partial cds.
ACCESSION   AF139445
VERSION     AF139445.1  GI:4928987
KEYWORDS    .
SOURCE      Gossypium hirsutum (upland cotton)
  ORGANISM  Gossypium hirsutum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium.
REFERENCE   1  (bases 1 to 1097)
  AUTHORS   Cronn,R.C., Small,R.L. and Wendel,J.F.
  TITLE     Duplicated genes evolve independently after polyploid formation in
            cotton
```

```
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (25), 14406-14411 (1999)
  MEDLINE   20056257
   PUBMED   10588718
REFERENCE   2  (bases 1 to 1097)
  AUTHORS   Cronn,R.C. and Wendel,J.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-MAR-1999) Botany, Iowa State University, 353 Bessey,
            Ames, IA 50011, USA
FEATURES             Location/Qualifiers
     source          1. .1097
                     /organism="Gossypium hirsutum"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3635"
                     /map="Homoeologous Assemblage 10"
                     /note="maps to the D-subgenome of G. hirsutum,
                     homoeologous assemblage 10; refer to Brubaker et al.
                     (Genome, 1999) for details."
     gene            <1. .>1097
                     /gene="CelA1"
     mRNA            join(<1. .151,325. .398,485. .536,630. .774,905. .>1097)
                     /gene="CelA1"
                     /product="cellulose synthase A1"
     CDS             join(<1. .151,325. .398,485. .536,630. .774,905. .>1097)
                     /gene="CelA1"
                     /codon_start=1
                     /product="cellulose synthase A1"
                     /protein_id="AAD33798.1"
                     /db_xref="GI:4928988"
                     /translation="HTCGEHVGLNVSGEPFVACHECNFPICESCFEYDLKEGRKACLR
                     CGSPYDENLLDDVEKATGDQSTMAAHLNKSQDVGIHARHISSVSTLDSEMAEDNGNSI
                     WKNRVESWKEKKNKKKKPATTKVERERAEIPPEQQMEDKPAPDASQPLSTIIPIPKSRL
                     APYRTVIIMRLIILGLFFHYRVTNPVDSAFGLWLTSVICEIWFAF"
BASE COUNT      320 a    180 c    239 g    358 t
ORIGIN

Alignment Scores:
Pred. No.:              3.85e-16        Length:         1097
Score:                  211.00          Matches:        33
Percent Similarity:     89.36%          Conservative:   9
Best Local Similarity:  70.21%          Mismatches:     5
Query Match:            65.12%          Indels:         0
DB:                     8               Gaps:           0

US-09-980-043-4 (1-56) x AF139445 (1-1097)

Qy         10 HisThrCysGlyGluGlnValGlyHisAspAlaAsnGlyGluLeuPheValAlaCysHis 29
              ||||||||||||||||   ||||  :::  :::||||  :||||||||||||||||
Db          1 CACACTTGTGGTGAACATGTAGGGTTGAATGTTAGTGGTGAACCTTTTGTGGCTTGCCAT 60

Qy         30 GluCysSerTyrProMetCysLysSerCysPheGluPheGluIleLysGluGlyArgLys 49
              |||||:::::|||:::||||:::|:||||.||||::::::::::|:||||||||||||||
Db         61 GAATGTAATTTCCCTATTTGTGAGAGTTGTTTTGAGTATGATCTTAAGGAAGGACGAAAA 120

Qy         50 ValCysLeuArgCysGlySer 56
              ||||||||||||||||||||
Db        121 GCTTGCTTGCGTTGTGGTAGT 141

RESULT 15
AF139449
LOCUS       AF139449               1201 bp    DNA     linear   PLN 12-JUL-2002
DEFINITION  Gossypium raimondii CelA2 pseudogene, partial sequence.
ACCESSION   AF139449
VERSION     AF139449.1  GI:4928993
KEYWORDS    .
SOURCE      Gossypium raimondii
  ORGANISM  Gossypium raimondii
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium.
REFERENCE   1  (bases 1 to 1201)
  AUTHORS   Cronn,R.C., Small,R.L. and Wendel,J.F.
  TITLE     Duplicated genes evolve independently after polyploid formation in
            cotton
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (25), 14406-14411 (1999)
  MEDLINE   20056257
   PUBMED   10588718
REFERENCE   2  (bases 1 to 1201)
  AUTHORS   Cronn,R.C. and Wendel,J.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-MAR-1999) Botany, Iowa State University, Bessey353,
            Ames, IA 50011, USA
FEATURES             Location/Qualifiers
     source          1. .1201
                     /organism="Gossypium raimondii"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:29730"
                     /map="Homoeologous assemblage 1"
                     /note="maps to homogologous assemblage 1; refer to
                     Brubaker et al. (Genome, 1999) for details."
     gene            <1. .>1201
                     /gene="CelA2"
                     /note="cellulose synthase A2"
                     /pseudo
     CDS             join(<1. .151,327. .400,521. .573,658. .799,1027. .>1201)
                     /gene="CelA2"
                     /pseudo
                     /codon_start=1
BASE COUNT      356 a    218 c    226 g    401 t
ORIGIN

Alignment Scores:
Pred. No.:              4.24e-16        Length:         1201
Score:                  211.00          Matches:        32
Percent Similarity:     89.36%          Conservative:   10
Best Local Similarity:  68.09%          Mismatches:     5
Query Match:            65.12%          Indels:         0
DB:                     8               Gaps:           0

US-09-980-043-4 (1-56) x AF139449 (1-1201)

Qy         10 HisThrCysGlyGluGlnValGlyHisAspAlaAsnGlyGluLeuPheValAlaCysHis 29
              ||||||||||||||||   ||||||   :::   |||||||||   ||||||||||||||||
Db          1 CACACCTGTGGTGAGCATGTTGGTTTGAATGGTAATGGAGAACCCTTCGTGGCTTGCCAT 60

Qy         30 GluCysSerTyrProMetCysLysSerCysPheGluPheGluIleLysGluGlyArgLys 49
              :::|||:::::::|||:::||||||:::||||||::::::|||:::||||||||||||||
Db         61 GATTGCAATTTCCCCATTTGCAAAACTTGTTTCGACTATGAACTTAAGGAAGGTCGAAAA 120

Qy         50 ValCysLeuArgCysGlySer 56
              ||||||||||||||||:::
Db        121 GCTTGTTTGCGTTGCGGTAAT 141


Search completed: October 31, 2003, 20:56:17
Job time : 621.731 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 29, 2003, 18:29:08 ; Search time 2.97872 Seconds
(without alignments)
3217.546 Million cell updates/sec

Title: US-09-980-043-4
Perfect score: 324
Sequence: 1 MMESGAPICHTCGEQVGHDA.....KSCFEFEIKEGRKVCLRCGS 56

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 642050 seqs, 171146064 residues

Total number of hits satisfying chosen parameters: 642050

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	264	81.5	974	10	US-09-838-539-6	Sequence 6, Appli
2	164	50.6	1091	10	US-09-900-237-26	Sequence 26, Appl
3	162	50.0	1075	12	US-10-209-059-2	Sequence 2, Appli
4	162	50.0	1075	12	US-10-160-719-10	Sequence 10, Appl
5	162	50.0	1075	12	US-10-160-719-34	Sequence 34, Appl
6	162	50.0	1075	12	US-10-160-719-54	Sequence 54, Appl
7	161	49.7	547	15	US-10-229-193-14	Sequence 14, Appl
8	161	49.7	1074	12	US-10-209-059-46	Sequence 46, Appl
9	161	49.7	1074	12	US-10-160-719-14	Sequence 14, Appl
10	161	49.7	1074	12	US-10-160-719-22	Sequence 22, Appl
11	161	49.7	1074	12	US-10-160-719-42	Sequence 42, Appl
12	158	48.8	1059	12	US-10-209-059-42	Sequence 42, Appl
13	158	48.8	1059	12	US-10-160-719-2	Sequence 2, Appli
14	158	48.8	1148	10	US-09-900-237-4	Sequence 4, Appli
15	154	47.5	1039	10	US-09-900-237-14	Sequence 14, Appl
16	153.5	47.4	1094	12	US-10-209-059-18	Sequence 18, Appl
17	153.5	47.4	1094	12	US-10-160-719-26	Sequence 26, Appl
18	153.5	47.4	1094	12	US-10-160-719-46	Sequence 46, Appl
19	153.5	47.4	1165	10	US-09-900-237-8	Sequence 8, Appli
20	152	46.9	1081	15	US-10-229-193-6	Sequence 6, Appli
21	152	46.9	1081	15	US-10-229-193-12	Sequence 12, Appl
22	150	46.3	1086	10	US-09-900-237-10	Sequence 10, Appl
23	150	46.3	1086	12	US-10-209-059-50	Sequence 50, Appl
24	150	46.3	1086	12	US-10-160-719-18	Sequence 18, Appl
25	150	46.3	1086	12	US-10-160-719-38	Sequence 38, Appl
26	148	45.7	1043	12	US-10-160-719-50	Sequence 50, Appl
27	148	45.7	1077	12	US-10-209-059-10	Sequence 10, Appl
28	148	45.7	1077	12	US-10-160-719-6	Sequence 6, Appli
29	148	45.7	1077	12	US-10-160-719-30	Sequence 30, Appl
30	148	45.7	1079	12	US-10-267-459-6	Sequence 6, Appli
31	148	45.7	1079	12	US-10-209-059-22	Sequence 22, Appl
32	148	45.7	1084	15	US-10-229-193-8	Sequence 8, Appli
33	146	45.1	1076	12	US-10-209-059-14	Sequence 14, Appl
34	146	45.1	1076	12	US-10-160-719-58	Sequence 58, Appl
35	138	42.6	1065	10	US-09-900-237-33	Sequence 33, Appl
36	138	42.6	1065	15	US-10-229-193-10	Sequence 10, Appl
37	138	42.6	1078	12	US-10-209-059-26	Sequence 26, Appl
38	135	41.7	1080	10	US-09-900-237-30	Sequence 30, Appl
39	115.5	35.6	1007	12	US-10-209-059-28	Sequence 28, Appl
40	77.5	23.9	592	11	US-09-919-039-141	Sequence 141, App
41	74	22.8	560	15	US-10-205-841-12	Sequence 12, Appl
42	72	22.2	168	9	US-09-864-761-43696	Sequence 43696, A
43	72	22.2	474	10	US-09-808-387-42	Sequence 42, Appl
44	72	22.2	474	12	US-10-442-754-2	Sequence 2, Appli
45	67.5	20.8	128	11	US-09-964-275B-10	Sequence 10, Appl

ALIGNMENTS

RESULT 1
US-09-838-539-6
; Sequence 6, Application US/09838539
; Patent No. US20020129401A1
; GENERAL INFORMATION:
; APPLICANT: Stalker, D. et al.
; TITLE OF INVENTION: Plant Cellulose Synthase and Promoter
; TITLE OF INVENTION: Sequences
; FILE REFERENCE: 15621/03/US
; CURRENT APPLICATION NUMBER: US/09/838,539
; CURRENT FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/029,987
; PRIOR FILING DATE: 1996-10-29
; PRIOR APPLICATION NUMBER: 08/960,048
; PRIOR FILING DATE: 1997-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 974
; TYPE: PRT
; ORGANISM: Gossypim hirsutum
US-09-838-539-6

Query Match 81.5%; Score 264; DB 10; Length 974;
Best Local Similarity 75.0%; Pred. No. 4.8e-22;
Matches 42; Conservative 8; Mismatches 6; Indels 0; Gaps 0;

Qy 1 MMESGAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEIKEGRKVCLRCGS 56
Db 1 MMESGVPVCHTCGEHVGLNVNGEPFVACHECNFPICKSCFEYDLKEGRKACLRCGS 56

RESULT 2
US-09-900-237-26
; Sequence 26, Application US/09900237
; Patent No. US20020120124A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Stephen

```
  TITLE OF INVENTION: Plant Cellulose Synthases
  FILE REFERENCE: BB1170 US CIP
  CURRENT APPLICATION NUMBER: US/09/900,237
  CURRENT FILING DATE:  2001-07-06
  PRIOR APPLICATION NUMBER: 60/092,844
  PRIOR FILING DATE: 1998-07-14
  PRIOR APPLICATION NUMBER: PCT/US99/15871
  PRIOR FILING DATE: 1999-07-13
  PRIOR APPLICATION NUMBER: 09/720383
  PRIOR FILING DATE: 2000-12-21
  NUMBER OF SEQ ID NOS: 33
  SOFTWARE: Microsoft Office 97
 SEQ ID NO 26
   LENGTH: 1091
   TYPE: PRT
   ORGANISM: Impatiens balsamia
JS-09-900-237-26

 Query Match            50.6%;  Score 164;  DB 10;  Length 1091;
 Best Local Similarity  44.4%;  Pred. No. 1.6e-10;
 Matches   24;  Conservative   13;  Mismatches   17;  Indels    0;  Gaps    0;

Qy          1 MMESGAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEIKEGRKVCLRC 54
              : |:    ||  ||: ||  | |: ||||:|| :|:|: |:|:| |:| : | :|
Db         33 LKEANGQICQICGDTVGKSATGDTFVACNECGFPVCRPCYEYERKDGNQCCPQC 86


ESULT 3
JS-10-209-059-2
 Sequence 2, Application US/10209059
 Publication No. US20030163838A1
 GENERAL INFORMATION:
  APPLICANT: Dhugga, Kanwarpal S.
  APPLICANT:  Wang, Haiyin
  TITLE OF INVENTION: Maize Cellulose Synthases and Uses
  TITLE OF INVENTION:  Thereof
  FILE REFERENCE: 0864R2
  CURRENT APPLICATION NUMBER: US/10/209,059
  CURRENT FILING DATE:  2002-07-31
  PRIOR APPLICATION NUMBER: 60/096,822
  PRIOR FILING DATE: 1998-08-17
  PRIOR APPLICATION NUMBER: 09/371,383
  PRIOR FILING DATE: 1999-08-06
  PRIOR APPLICATION NUMBER: 09/550,483
  PRIOR FILING DATE: 2000-04-14
  NUMBER OF SEQ ID NOS: 52
  SOFTWARE: FastSEQ for Windows Version 4.0
 SEQ ID NO 2
   LENGTH: 1075
   TYPE: PRT
   ORGANISM: Zea mays
JS-10-209-059-2

 Query Match            50.0%;  Score 162;  DB 12;  Length 1075;
 Best Local Similarity  48.9%;  Pred. No. 2.7e-10;
 Matches   23;  Conservative   13;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          8 ICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEIKEGRKVCLRC 54
              :|  ||: ||  | |::||||:||::|:|: |:|:| ||| : | :|
Db         40 VCQICGDSVGVSATGDVFVACNECAFPVCRPCYEYERKEGNQCCPQC 86


RESULT 4
JS-10-160-719-10
 Sequence 10, Application US/10160719
 Publication No. US20030167528A1
 GENERAL INFORMATION:
  APPLICANT: Kanwarpal S. Dhugga
  APPLICANT:  Timothy G. Helentjaris
  APPLICANT:  Benjamin A.  Bowen
  APPLICANT:  Xun Wang
  TITLE OF INVENTION: Maize Cellulose Synthases and Uses
  TITLE OF INVENTION:  Thereof
;  FILE REFERENCE: 0864
;  CURRENT APPLICATION NUMBER: US/10/160,719
;  CURRENT FILING DATE:  2002-06-03
;  PRIOR APPLICATION NUMBER: US/09/371,383A
;  PRIOR FILING DATE: 1999-08-06
;  PRIOR APPLICATION NUMBER: US 60/096,822
;  PRIOR FILING DATE: 1998-08-17
;  NUMBER OF SEQ ID NOS: 60
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 10
;   LENGTH: 1075
;   TYPE: PRT
;   ORGANISM: Zea mays
US-10-160-719-10

  Query Match            50.0%;  Score 162;  DB 12;  Length 1075;
  Best Local Similarity  48.9%;  Pred. No. 2.7e-10;
  Matches   23;  Conservative   13;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          8 ICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEIKEGRKVCLRC 54
              :|  ||: ||  | |::||||:||::|:|: |:|:| ||| : | :|
Db         40 VCQICGDSVGVSATGDVFVACNECAFPVCRPCYEYERKEGNQCCPQC 86


RESULT 5
US-10-160-719-34
; Sequence 34, Application US/10160719
; Publication No. US20030167528A1
; GENERAL INFORMATION:
;  APPLICANT: Kanwarpal S. Dhugga
;  APPLICANT:  Timothy G. Helentjaris
;  APPLICANT:  Benjamin A.  Bowen
;  APPLICANT:  Xun Wang
;  TITLE OF INVENTION: Maize Cellulose Synthases and Uses
;  TITLE OF INVENTION:  Thereof
;  FILE REFERENCE: 0864
;  CURRENT APPLICATION NUMBER: US/10/160,719
;  CURRENT FILING DATE:  2002-06-03
;  PRIOR APPLICATION NUMBER: US/09/371,383A
;  PRIOR FILING DATE: 1999-08-06
;  PRIOR APPLICATION NUMBER: US 60/096,822
;  PRIOR FILING DATE: 1998-08-17
;  NUMBER OF SEQ ID NOS: 60
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 34
;   LENGTH: 1075
;   TYPE: PRT
;   ORGANISM: Zea mays
US-10-160-719-34

  Query Match            50.0%;  Score 162;  DB 12;  Length 1075;
  Best Local Similarity  48.9%;  Pred. No. 2.7e-10;
  Matches   23;  Conservative   13;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          8 ICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEIKEGRKVCLRC 54
              :|  ||: ||  | |::||||:||::|:|: |:|:| ||| : | :|
Db         40 VCQICGDSVGVSATGDVFVACNECAFPVCRPCYEYERKEGNQCCPQC 86


RESULT 6
US-10-160-719-54
; Sequence 54, Application US/10160719
; Publication No. US20030167528A1
; GENERAL INFORMATION:
;  APPLICANT: Kanwarpal S. Dhugga
;  APPLICANT:  Timothy G. Helentjaris
;  APPLICANT:  Benjamin A.  Bowen
;  APPLICANT:  Xun Wang
;  TITLE OF INVENTION: Maize Cellulose Synthases and Uses
;  TITLE OF INVENTION:  Thereof
;  FILE REFERENCE: 0864
;  CURRENT APPLICATION NUMBER: US/10/160,719
;  CURRENT FILING DATE:  2002-06-03
```

; PRIOR APPLICATION NUMBER: US/09/371,383A
; PRIOR FILING DATE: 1999-08-06
; PRIOR APPLICATION NUMBER: US 60/096,822
; PRIOR FILING DATE: 1998-08-17
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 54
; LENGTH: 1075
; TYPE: PRT
; ORGANISM: Zea mays
US-10-160-719-54

Query Match 50.0%; Score 162; DB 12; Length 1075;
Best Local Similarity 48.9%; Pred. No. 2.7e-10;
Matches 23; Conservative 13; Mismatches 11; Indels 0; Gaps 0;

Qy 8 ICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEIKEGRKVCLRC 54
Db 40 VCQICGDSVGVSATGDVFVACNECAFPVCRPCYEYERKEGNQCCPQC 86

RESULT 7
US-10-229-193-14
; Sequence 14, Application US/10229193
; Publication No. US20030126643A1
; GENERAL INFORMATION:
; APPLICANT: Arioli, Antonio Williamson, Richard E.
; APPLICANT: Betzner, Andreas S.
; APPLICANT: Peng, Liangcai
; TITLE OF INVENTION: Manipulation of Cellulose and/or Beta 1.4-Glucan
; FILE REFERENCE: 96-98A
; CURRENT APPLICATION NUMBER: US/10/229,193
; CURRENT FILING DATE: 2002-08-26
; PRIOR APPLICATION NUMBER: 09/221,013
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: PCT/AU97/00402
; PRIOR FILING DATE: 1997-06-24
; PRIOR APPLICATION NUMBER: AU PO0699
; PRIOR FILING DATE: 1996-06-27
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
; LENGTH: 547
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-229-193-14

Query Match 49.7%; Score 161; DB 12; Length 547;
Best Local Similarity 48.9%; Pred. No. 1.2e-10;
Matches 23; Conservative 13; Mismatches 11; Indels 0; Gaps 0;

Qy 8 ICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEIKEGRKVCLRC 54
Db 40 VCQICGDTVGVSATGDVFVACNECAFPVCRPCYEYERKEGNQCCPQC 86

RESULT 8
US-10-209-059-46
; Sequence 46, Application US/10209059
; Publication No. US20030163838A1
; GENERAL INFORMATION:
; APPLICANT: Dhugga, Kanwarpal S.
; APPLICANT: Wang, Haiyin
; TITLE OF INVENTION: Maize Cellulose Synthases and Uses
; TITLE OF INVENTION: Thereof
; FILE REFERENCE: 0864R2
; CURRENT APPLICATION NUMBER: US/10/209,059
; CURRENT FILING DATE: 2002-07-31
; PRIOR APPLICATION NUMBER: 60/096,822
; PRIOR FILING DATE: 1998-08-17
; PRIOR APPLICATION NUMBER: 09/371,383
; PRIOR FILING DATE: 1999-08-06
; PRIOR APPLICATION NUMBER: 09/550,483
; PRIOR FILING DATE: 2000-04-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 46
; LENGTH: 1074
; TYPE: PRT
; ORGANISM: Zea mays
US-10-209-059-46

Query Match 49.7%; Score 161; DB 12; Length 1074;
Best Local Similarity 48.9%; Pred. No. 3.5e-10;
Matches 23; Conservative 13; Mismatches 11; Indels 0; Gaps 0;

Qy 8 ICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEIKEGRKVCLRC 54
Db 40 VCQICGDTVGVSATGDVFVACNECAFPVCRPCYEYERKEGNQCCPQC 86

RESULT 9
US-10-160-719-14
; Sequence 14, Application US/10160719
; Publication No. US20030167528A1
; GENERAL INFORMATION:
; APPLICANT: Kanwarpal S. Dhugga
; APPLICANT: Timothy G. Helentjaris
; APPLICANT: Benjamin A. Bowen
; APPLICANT: Xun Wang
; TITLE OF INVENTION: Maize Cellulose Synthases and Uses
; TITLE OF INVENTION: Thereof
; FILE REFERENCE: 0864
; CURRENT APPLICATION NUMBER: US/10/160,719
; CURRENT FILING DATE: 2002-06-03
; PRIOR APPLICATION NUMBER: US/09/371,383A
; PRIOR FILING DATE: 1999-08-06
; PRIOR APPLICATION NUMBER: US 60/096,822
; PRIOR FILING DATE: 1998-08-17
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 14
; LENGTH: 1074
; TYPE: PRT
; ORGANISM: Zea mays
US-10-160-719-14

Query Match 49.7%; Score 161; DB 12; Length 1074;
Best Local Similarity 48.9%; Pred. No. 3.5e-10;
Matches 23; Conservative 13; Mismatches 11; Indels 0; Gaps 0;

Qy 8 ICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEIKEGRKVCLRC 54
Db 40 VCQICGDTVGVSATGDVFVACNECAFPVCRPCYEYERKEGNQCCPQC 86

RESULT 10
US-10-160-719-22
; Sequence 22, Application US/10160719
; Publication No. US20030167528A1
; GENERAL INFORMATION:
; APPLICANT: Kanwarpal S. Dhugga
; APPLICANT: Timothy G. Helentjaris
; APPLICANT: Benjamin A. Bowen
; APPLICANT: Xun Wang
; TITLE OF INVENTION: Maize Cellulose Synthases and Uses
; TITLE OF INVENTION: Thereof
; FILE REFERENCE: 0864
; CURRENT APPLICATION NUMBER: US/10/160,719
; CURRENT FILING DATE: 2002-06-03
; PRIOR APPLICATION NUMBER: US/09/371,383A
; PRIOR FILING DATE: 1999-08-06
; PRIOR APPLICATION NUMBER: US 60/096,822
; PRIOR FILING DATE: 1998-08-17
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: FastSEQ for Windows Version 3.0

```
 SEQ ID NO 22
   LENGTH: 1074
   TYPE: PRT
   ORGANISM: Zea mays
S-10-160-719-22

 Query Match            49.7%;  Score 161;  DB 12;  Length 1074;
 Best Local Similarity  48.9%;  Pred. No. 3.5e-10;
 Matches   23;  Conservative   13;  Mismatches   11;  Indels    0;  Gaps    0;

y          8 ICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEIKEGRKVCLRC 54
             :|  ||: ||  | |::||||:||::|:|: |:|:| ||| : | :|
b         40 VCQICGDTVGVSATGDVFVACNECAFPVCRPCYEYERKEGNQCCPQC 86


ESULT 11
S-10-160-719-42
 Sequence 42, Application US/10160719
 Publication No. US20030167528A1
 GENERAL INFORMATION:
  APPLICANT: Kanwarpal S. Dhugga
  APPLICANT:  Timothy G. Helentjaris
  APPLICANT:  Benjamin A.  Bowen
  APPLICANT:  Xun Wang
  TITLE OF INVENTION: Maize Cellulose Synthases and Uses
  TITLE OF INVENTION:  Thereof
  FILE REFERENCE: 0864
  CURRENT APPLICATION NUMBER: US/10/160,719
  CURRENT FILING DATE:  2002-06-03
  PRIOR APPLICATION NUMBER: US/09/371,383A
  PRIOR FILING DATE: 1999-08-06
  PRIOR APPLICATION NUMBER: US 60/096,822
  PRIOR FILING DATE: 1998-08-17
  NUMBER OF SEQ ID NOS: 60
  SOFTWARE: FastSEQ for Windows Version 3.0
 SEQ ID NO 42
   LENGTH: 1074
   TYPE: PRT
   ORGANISM: Zea mays
JS-10-160-719-42

 Query Match            49.7%;  Score 161;  DB 12;  Length 1074;
 Best Local Similarity  48.9%;  Pred. No. 3.5e-10;
 Matches   23;  Conservative   13;  Mismatches   11;  Indels    0;  Gaps    0;

Qy         8 ICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEIKEGRKVCLRC 54
             :|  ||: ||  | |::||||:||::|:|: |:|:| ||| : | :|
Db        40 VCQICGDTVGVSATGDVFVACNECAFPVCRPCYEYERKEGNQCCPQC 86


ESULT 12
S-10-209-059-42
 Sequence 42, Application US/10209059
 Publication No. US20030163838A1
 GENERAL INFORMATION:
  APPLICANT: Dhugga, Kanwarpal S.
  APPLICANT:  Wang, Haiyin
  TITLE OF INVENTION: Maize Cellulose Synthases and Uses
  TITLE OF INVENTION:  Thereof
  FILE REFERENCE: 0864R2
  CURRENT APPLICATION NUMBER: US/10/209,059
  CURRENT FILING DATE:  2002-07-31
  PRIOR APPLICATION NUMBER: 60/096,822
  PRIOR FILING DATE: 1998-08-17
  PRIOR APPLICATION NUMBER: 09/371,383
  PRIOR FILING DATE: 1999-08-06
  PRIOR APPLICATION NUMBER: 09/550,483
  PRIOR FILING DATE: 2000-04-14
  NUMBER OF SEQ ID NOS: 52
  SOFTWARE: FastSEQ for Windows Version 4.0
 SEQ ID NO 42
   LENGTH: 1059
   TYPE: PRT
;   ORGANISM: Zea mays
US-10-209-059-42

  Query Match            48.8%;  Score 158;  DB 12;  Length 1059;
  Best Local Similarity  42.6%;  Pred. No. 7.6e-10;
  Matches   23;  Conservative   15;  Mismatches   16;  Indels    0;  Gaps    0;

Qy         1 MMESGAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEIKEGRKVCLRC 54
             |  :    :|  ||: '| : :||  .||:|:::|:|: |:|:| :|| : | :|
Db         1 MDQRNGQVCQICGDEVGRNPDGEPFVACNECAFPICRDCYEYERREGTQNCPQC 54


RESULT 13
US-10-160-719-2
; Sequence 2, Application US/10160719
; Publication No. US20030167528A1
; GENERAL INFORMATION:
;  APPLICANT: Kanwarpal S. Dhugga
;  APPLICANT:  Timothy G. Helentjaris
;  APPLICANT:  Benjamin A.  Bowen
;  APPLICANT:  Xun Wang
;  TITLE OF INVENTION: Maize Cellulose Synthases and Uses
;  TITLE OF INVENTION:  Thereof
;  FILE REFERENCE: 0864
;  CURRENT APPLICATION NUMBER: US/10/160,719
;  CURRENT FILING DATE:  2002-06-03
;  PRIOR APPLICATION NUMBER: US/09/371,383A
;  PRIOR FILING DATE: 1999-08-06
;  PRIOR APPLICATION NUMBER: US 60/096,822
;  PRIOR FILING DATE: 1998-08-17
;  NUMBER OF SEQ ID NOS: 60
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
;   LENGTH: 1059
;   TYPE: PRT
;   ORGANISM: Zea mays
US-10-160-719-2

  Query Match            48.8%;  Score 158;  DB 12;  Length 1059;
  Best Local Similarity  42.6%;  Pred. No. 7.6e-10;
  Matches   23;  Conservative   15;  Mismatches   16;  Indels    0;  Gaps    0;

Qy         1 MMESGAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEIKEGRKVCLRC 54
             |  :    :|  ||: || : :|| ||||:.'::|::: |:|:| :|| : | :|
Db         1 MDQRNGQVCQICGDEVGRNPDGEPFVACNECAFPICRDCYEYERREGTQNCPQC 54


RESULT 14
US-09-900-237-4
; Sequence 4, Application US/09900237
; Patent No. US20020120124A1
; GENERAL INFORMATION:
;  APPLICANT: Allen, Stephen
;  TITLE OF INVENTION: Plant Cellulose Synthases
;  FILE REFERENCE: BB1170 US CIP
;  CURRENT APPLICATION NUMBER: US/09/900,237
;  CURRENT FILING DATE:  2001-07-06
;  PRIOR APPLICATION NUMBER: 60/092,844
;  PRIOR FILING DATE: 1998-07-14
;  PRIOR APPLICATION NUMBER: PCT/US99/15871
;  PRIOR FILING DATE: 1999-07-13
;  PRIOR APPLICATION NUMBER: 09/720383
;  PRIOR FILING DATE: 2000-12-21
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 4
;   LENGTH: 1148
;   TYPE: PRT
;   ORGANISM: Zea mays
US-09-900-237-4

  Query Match            48.8%;  Score 158;  DB 10;  Length 1148;
  Best Local Similarity  42.6%;  Pred. No. 8.3e-10;
```

Matches 23; Conservative 15; Mismatches 16; Indels 0; Gaps 0;

```
Qy          1 MMESGAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEIKEGRKVCLRC 54
              |  :    :|  ||: || : :|| ||||:||::|:|: |:|:| :|| : | :|
Db         90 MDQRNGQVCQICGDDVGRNPDGEPFVACNECAFPICRDCYEYERREGTQNCPQC 143
```

RESULT 15
US-09-900-237-14
; Sequence 14, Application US/09900237
; Patent No. US20020120124A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Stephen
; TITLE OF INVENTION: Plant Cellulose Synthases
; FILE REFERENCE: BB1170 US CIP
; CURRENT APPLICATION NUMBER: US/09/900,237
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: 60/092,844
; PRIOR FILING DATE: 1998-07-14
; PRIOR APPLICATION NUMBER: PCT/US99/15871
; PRIOR FILING DATE: 1999-07-13
; PRIOR APPLICATION NUMBER: 09/720383
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 14
; LENGTH: 1039
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (201)
US-09-900-237-14

Query Match 47.5%; Score 154; DB 10; Length 1039;
Best Local Similarity 46.8%; Pred. No. 2.1e-09;
Matches 22; Conservative 13; Mismatches 12; Indels 0; Gaps 0;

```
Qy          8 ICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEIKEGRKVCLRC 54
              :|  ||: ||   :|:|||':|| :|:|: |:|:| :||  :| :|
Db         36 VCEICGDGVGLTVDGDLFVACNECGFPVCRPCYEYERREGSHLCPQC 82
```

Search completed: October 29, 2003, 18:33:07
Job time : 3.97872 secs

GenCore version 5.1.6
Copyright (c) 1993 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 29, 2003, 18:27:33 ; Search time 1.62476 Seconds
(without alignments)
1458.314 Million cell updates/sec

Title: US-09-980-043-4
Perfect score: 324
Sequence: 1 MMESGAPICHTCGEQVGHDA.....KSCFEFEIKEGRKVCLRCGS 56

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	264	81.5	974	3	US-09-960-048-6	Sequence 6, Appli
2	264	81.5	974	4	US-09-838-586-6	Sequence 6, Appli
3	161	49.7	547	4	US-09-221-013A-14	Sequence 14, Appl
4	152	46.9	1081	4	US-09-221-013A-6	Sequence 6, Appli
5	152	46.9	1081	4	US-09-221-013A-12	Sequence 12, Appl
6	148	45.7	1084	4	US-09-221-013A-8	Sequence 8, Appli
7	138	42.6	1065	4	US-09-221-013A-10	Sequence 10, Appl
8	67	20.7	1400	1	US-08-080-255-7	Sequence 7, Appli
9	67	20.7	1400	3	US-08-465-713-7	Sequence 7, Appli
10	67	20.7	1400	5	PCT-US93-05857-7	Sequence 7, Appli
11	67	20.7	3969	3	US-08-061-376-5	Sequence 5, Appli
12	66	20.4	647	2	US-08-678-039A-42	Sequence 42, Appl
13	65.5	20.2	124	3	US-08-946-026-31	Sequence 31, Appl
14	65.5	20.2	314	1	US-08-269-441A-2	Sequence 2, Appli
15	65.5	20.2	314	3	US-09-008-465-3	Sequence 3, Appli
16	65.5	20.2	314	3	US-09-276-861-2	Sequence 2, Appli
17	65.5	20.2	314	4	US-09-528-969-3	Sequence 3, Appli
18	65.5	20.2	739	4	US-09-252-991A-32778	Sequence 32778, A
19	65	20.1	265	2	US-08-937-931-6	Sequence 6, Appli
20	65	20.1	265	3	US-09-285-502-6	Sequence 6, Appli
21	65	20.1	265	4	US-09-709-126-6	Sequence 6, Appli
22	65	20.1	265	4	US-09-871-385A-6	Sequence 6, Appli
23	65	20.1	491	3	US-09-030-335-2	Sequence 2, Appli
24	65	20.1	748	2	US-08-920-234-2	Sequence 2, Appli
25	65	20.1	748	2	US-08-937-931-4	Sequence 4, Appli
26	65	20.1	748	3	US-09-285-502-4	Sequence 4, Appli
27	65	20.1	748	3	US-09-030-335-9	Sequence 9, Appli
28	65	20.1	748	4	US-09-709-126-4	Sequence 4, Appli
29	65	20.1	748	4	US-09-871-385A-4	Sequence 4, Appli
30	65	20.1	799	3	US-09-030-335-4	Sequence 4, Appli
31	63	19.4	738	3	US-08-478-208-32	Sequence 32, Appl
32	63	19.4	738	4	US-09-336-536-73	Sequence 73, Appl
33	63	19.4	738	6	5264554-2	Patent No. 5264554
34	62	19.1	366	3	US-08-945-056-6	Sequence 6, Appli
35	62	19.1	749	2	US-08-937-931-8	Sequence 8, Appli
36	62	19.1	749	3	US-09-285-502-8	Sequence 8, Appli
37	62	19.1	749	4	US-09-709-126-8	Sequence 8, Appli
38	62	19.1	749	4	US-09-871-385A-8	Sequence 8, Appli
39	61.5	19.0	1940	2	US-08-644-271-30	Sequence 30, Appl
40	61.5	19.0	1940	4	US-09-077-955-34	Sequence 34, Appl
41	61	18.8	366	3	US-08-945-056-8	Sequence 8, Appli
42	61	18.8	453	4	US-09-686-583B-12	Sequence 12, Appl
43	60.5	18.7	113	1	US-08-320-559-37	Sequence 37, Appl
44	60.5	18.7	113	3	US-08-545-860D-37	Sequence 37, Appl
45	60.5	18.7	113	5	PCT-US94-04496-37	Sequence 37, Appl

ALIGNMENTS

RESULT 1
US-08-960-048-6
; Sequence 6, Application US/08960048C
; Patent No. 6271443
; GENERAL INFORMATION:
; APPLICANT: Stalker, D. et al.
; TITLE OF INVENTION: Plant Cellulose Synthase and Promoter
; TITLE OF INVENTION: Sequences
; FILE REFERENCE: 15621/01/US
; CURRENT APPLICATION NUMBER: US/08/960,048C
; CURRENT FILING DATE: 1997-10-29
; PRIOR APPLICATION NUMBER: 60/029,987
; PRIOR FILING DATE: 1996-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 974
; TYPE: PRT
; ORGANISM: Gossypim hirsutum
US-08-960-048-6

Query Match 81.5%; Score 264; DB 3; Length 974;
Best Local Similarity 75.0%; Pred. No. 8.1e-23;
Matches 42; Conservative 8; Mismatches 6; Indels 0; Gaps 0;

Qy 1 MMESGAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEIKEGRKVCLRCGS 56
Db 1 MMESGVPVCHTCGEHVGLNVNGEPFVACHECNFPICKSCFEYDLKEGRKACLRCGS 56

RESULT 2
US-09-838-586-6
; Sequence 6, Application US/09838586
; Patent No. 6576818
; GENERAL INFORMATION:
; APPLICANT: Stalker, D. et al.
; TITLE OF INVENTION: Plant Cellulose Synthase and Promoter
; TITLE OF INVENTION: Sequences
; FILE REFERENCE: 15621/02/US
; CURRENT APPLICATION NUMBER: US/09/838,586
; CURRENT FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/029,987
; PRIOR FILING DATE: 1996-10-29
; PRIOR APPLICATION NUMBER: 08/960,048
; PRIOR FILING DATE: 1997-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 974
; TYPE: PRT

ORGANISM: Gossypim hirsutum
US-09-838-586-6

Query Match 81.5%; Score 264; DB 4; Length 974;
Best Local Similarity 75.0%; Pred. No. 8.1e-23;
Matches 42; Conservative 8; Mismatches 6; Indels 0; Gaps 0;

Qy 1 MMESGAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEIKEGRKVCLRCGS 56
Db 1 MMESGVPVCHTCGEHVGLNVNGEPFVACHECNFPICKSCFEYDLKEGRKACLRCGS 56

RESULT 3
US-09-221-013A-14
Sequence 14, Application US/09221013A
Patent No. 6495740
GENERAL INFORMATION:
APPLICANT: Arioli, Antonio
APPLICANT: Williamson, Richard E.
APPLICANT: Betzner, Andreas S.
APPLICANT: Peng, Liangcai
TITLE OF INVENTION: Manipulation of cellulose and/or Beta-1,4-glucan
FILE REFERENCE: 96-98
CURRENT APPLICATION NUMBER: US/09/221,013A
CURRENT FILING DATE: 1998-12-23
PRIOR APPLICATION NUMBER: PCT/AU97/00402
PRIOR FILING DATE: 1997-06-24
PRIOR APPLICATION NUMBER: AU PO0699
PRIOR FILING DATE: 1996-06-27
NUMBER OF SEQ ID NOS: 37
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 14
LENGTH: 547
TYPE: PRT
ORGANISM: Oryza sativa
US-09-221-013A-14

Query Match 49.7%; Score 161; DB 4; Length 547;
Best Local Similarity 48.9%; Pred. No. 4.4e-11;
Matches 23; Conservative 13; Mismatches 11; Indels 0; Gaps 0;

Qy 8 ICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEIKEGRKVCLRC 54
Db 40 VCQICGDTVGVSATGDVFVACNECAFPVCRPCYEYERKEGNQCCPQC 86

RESULT 4
US-09-221-013A-6
Sequence 6, Application US/09221013A
Patent No. 6495740
GENERAL INFORMATION:
APPLICANT: Arioli, Antonio
APPLICANT: Williamson, Richard E.
APPLICANT: Betzner, Andreas S.
APPLICANT: Peng, Liangcai
TITLE OF INVENTION: Manipulation of cellulose and/or Beta-1,4-glucan
FILE REFERENCE: 96-98
CURRENT APPLICATION NUMBER: US/09/221,013A
CURRENT FILING DATE: 1998-12-23
PRIOR APPLICATION NUMBER: PCT/AU97/00402
PRIOR FILING DATE: 1997-06-24
PRIOR APPLICATION NUMBER: AU PO0699
PRIOR FILING DATE: 1996-06-27
NUMBER OF SEQ ID NOS: 37
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 6
LENGTH: 1081
TYPE: PRT
ORGANISM: Arabidopsis thaliana
US-09-221-013A-6

Query Match 46.9%; Score 152; DB 4; Length 1081;
Best Local Similarity 46.8%; Pred. No. 1e-09;
Matches 22; Conservative 13; Mismatches 12; Indels 0; Gaps 0;

Qy 8 ICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEIKEGRKVCLRC 54
Db 38 ICQICGDDVGLAETGDVFVACNECAFPVCRPCYEYERKDGTQCCPQC 84

RESULT 5
US-09-221-013A-12
; Sequence 12, Application US/09221013A
; Patent No. 6495740
; GENERAL INFORMATION:
; APPLICANT: Arioli, Antonio
; APPLICANT: Williamson, Richard E
; APPLICANT: Betzner, Andreas S
; APPLICANT: Peng, Liangcai
; TITLE OF INVENTION: Manipulation of cellulose and/or Beta-1,4-glucan
; FILE REFERENCE: 96-98
; CURRENT APPLICATION NUMBER: US/09221013A
; CURRENT FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: PCT/AU97/00402
; PRIOR FILING DATE: 1997-06-24
; PRIOR APPLICATION NUMBER: AU PO0699
; PRIOR FILING DATE: 1996-06-27
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 1081
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-221-013A-12

Query Match 46.9%; Score 152; DB 4; Length 1081;
Best Local Similarity 46.8%; Pred. No. 1e-09;
Matches 22; Conservative 13; Mismatches 12; Indels 0; Gaps 0;

Qy 8 ICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEIKEGRKVCLRC 54
Db 38 ICQICGDDVGLAETGDVFVACNECAFPVCRPCYEYERKDGTQCCPQC 84

RESULT 6
US-09-221-013A-8
; Sequence 8, Application US/09221013A
; Patent No. 6495740
; GENERAL INFORMATION:
; APPLICANT: Arioli, Antonio
; APPLICANT: Williamson, Richard E.
; APPLICANT: Betzner, Andreas S.
; APPLICANT: Peng, Liangcai
; TITLE OF INVENTION: Manipulation of cellulose and/or Beta-1,4-glucan
; FILE REFERENCE: 96-98
; CURRENT APPLICATION NUMBER: US/09/221,013A
; CURRENT FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: PCT/AU97/00402
; PRIOR FILING DATE: 1997-06-24
; PRIOR APPLICATION NUMBER: AU PO0699
; PRIOR FILING DATE: 1996-06-27
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 1084
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-221-013A-8

Query Match 45.7%; Score 148; DB 4; Length 1084;
Best Local Similarity 38.9%; Pred. No. 2.9e-09;
Matches 21; Conservative 15; Mismatches 18; Indels 0; Gaps 0;

Qy 1 MMESGAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEIKEGRKVCLRC 54
Db 31 VQELSGQTCQICGDEIELTVSSELFVACNECAFPVCRPCYEYERREGNQACPQC 84

```
RESULT 7
US-09-221-013A-10
; Sequence 10, Application US/09221013A
; Patent No. 6495740
; GENERAL INFORMATION:
;  APPLICANT:  Arioli, Antonio
;  APPLICANT:  Williamson, Richard E.
;  APPLICANT:  Betzner, Andreas S.
;  APPLICANT:  Peng, Liangcai
;  TITLE OF INVENTION:  Manipulation of cellulose and/or Beta-1,4-glucan
;  FILE REFERENCE:  96-98
;  CURRENT APPLICATION NUMBER: US/09/221,013A
;  CURRENT FILING DATE:  1998-12-23
;  PRIOR APPLICATION NUMBER:  PCT/AU97/00402
;  PRIOR FILING DATE:  1997-06-24
;  PRIOR APPLICATION NUMBER:  AU PO0699
;  PRIOR FILING DATE:  1996-06-27
;  NUMBER OF SEQ ID NOS:  37
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 10
;   LENGTH: 1065
;   TYPE: PRT
;   ORGANISM: Arabidopsis thaliana
US-09-221-013A-10

  Query Match             42.6%;  Score 138;  DB 4;  Length 1065;
  Best Local Similarity   38.5%;  Pred. No. 4.2e-08;
  Matches   25;  Conservative   12;  Mismatches   16;  Indels   12;  Gaps    2;

Qy          2 MES-----GAPI-------CHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEIKEGRK 49
              |||     | |:       |  | : :|   :|: |||   |:|:|:|: |:|:| |:| :
Db          1 MESEGETAGKPMKNIVPQTCQICSDNVGKTVDGDRFVACDICSFPVCRPCYEYERKDGNQ 60

Qy         50 VCLRC 54
                | :|
Db         61 SCPQC 65


RESULT 8
US-08-080-255-7
; Sequence 7, Application US/08080255
; Patent No. 5487970
;  GENERAL INFORMATION:
;    APPLICANT:  Rowley, Janet D.
;    APPLICANT:  Diaz, Manuel O.
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR
;    TITLE OF INVENTION:  DETECTING GENE REARRANGEMENTS AND
;    TITLE OF INVENTION:  TRANSLOCATIONS
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P. O. Box 4433
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  USA
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/080,255
;      FILING DATE:  19930617
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Parker, David L.
;      REGISTRATION NUMBER:  32,165
;      REFERENCE/DOCKET NUMBER:  ARCD:072/PAR
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (512) 320-7200
;      TELEFAX:  (512) 474-7577
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1400 amino acids
;      TYPE:  AMINO ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-080-255-7

  Query Match             20.7%;  Score 67;  DB 1;  Length 1400;
  Best Local Similarity   25.0%;  Pred. No. 10;
  Matches   16;  Conservative   10;  Mismatches   14;  Indels   24;  Gaps    4;

Qy          9 CHTCGE---------QVGHDANGELFVACHECS--------YPMCKSCFEFEIKEGRKVC 51
              | :||          |  ||     |  ||:|:         |:|  |:: :  | :
Db        718 CKSCGSTTPGKGWDAQWSHD-----FSLCHDCAKLFAKGNFCPLCDKCYDDDDYESK--M 770

Qy         52 LRCG 55
                ::||
Db        771 MQCG 774


RESULT 9
US-08-465-713-7
; Sequence 7, Application US/08465713
; Patent No. 6121419
;  GENERAL INFORMATION:
;    APPLICANT:  Rowley, Janet D.
;    APPLICANT:  Diaz, Manuel O.
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR
;    TITLE OF INVENTION:  DETECTING GENE REARRANGEMENTS AND
;    TITLE OF INVENTION:  TRANSLOCATIONS
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P. O. Box 4433
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  USA
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/465,713
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/080,255
;      FILING DATE:  17 JUNE 1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Parker, David L.
;      REGISTRATION NUMBER:  32,165
;      REFERENCE/DOCKET NUMBER:  ARCD:072/PAR
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (512) 320-7200
;      TELEFAX:  (512) 474-7577
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1400 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-465-713-7

  Query Match             20.7%;  Score 67;  DB 3;  Length 1400;
  Best Local Similarity   25.0%;  Pred. No. 10;
  Matches   16;  Conservative   10;  Mismatches   14;  Indels   24;  Gaps    4;
```

```
Qy          9 CHTCGE---------QVGHDANGELFVACHECS--------YPMCKSCFEFEIKEGRKVC 51
              | :||           |  ||      |  ||:|:          |:|  |:: :  | :
Db        718 CKSCGSTTPGKGWDAQWSHD-----FSLCHDCAKLFAKGNFCPLCDKCYDDDDYESK--M 770

Qy         52 LRCG 55
              ::||
Db        771 MQCG 774


RESULT 10
US-09-05857-7
 Sequence 7, Application PC/TUS9305857
  GENERAL INFORMATION:
    APPLICANT:  Board of Regents
    APPLICANT:  The University of Texas System
    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR DETECTING
    TITLE OF INVENTION:  GENE REARRANGEMENTS AND TRANSLOCATIONS
    NUMBER OF SEQUENCES:  8
    CORRESPONDENCE ADDRESS:
      ADDRESSEE:  Arnold, White & Durkee
      STREET:  P. O. Box 4433
      CITY:  Houston
      STATE:  Texas
      COUNTRY:  USA
      ZIP:  77210
    COMPUTER READABLE FORM:
      MEDIUM TYPE:  Floppy disk
      COMPUTER:  IBM PC compatible
      OPERATING SYSTEM:  PC-DOS/MS-DOS
      SOFTWARE:  PatentIn Release #1.0, Version #1.25
    CURRENT APPLICATION DATA:
      APPLICATION NUMBER:  PCT/US93/05857
      FILING DATE:  19930617
      CLASSIFICATION:
    PRIOR APPLICATION DATA:
      APPLICATION NUMBER:  07/900,689
      FILING DATE:  17/06/92
    ATTORNEY/AGENT INFORMATION:
      NAME:  Parker, David L.
      REGISTRATION NUMBER:  32,165
      REFERENCE/DOCKET NUMBER:  ARCD:072/PAR
    TELECOMMUNICATION INFORMATION:
      TELEPHONE:  (512) 320-7200
      TELEFAX:  (512) 474-7577
  INFORMATION FOR SEQ ID NO:  7:
    SEQUENCE CHARACTERISTICS:
      LENGTH:  1400 amino acids
      TYPE:  amino acid
      STRANDEDNESS:  single
      TOPOLOGY:  linear
    MOLECULE TYPE:  DNA (genomic)
US-09-05857-7

 Query Match             20.7%;  Score 67;  DB 5;  Length 1400;
 Best Local Similarity   25.0%;  Pred. No. 10;
 Matches   16;  Conservative   10;  Mismatches   14;  Indels   24;  Gaps    4;

Qy          9 CHTCGE---------QVGHDANGELFVACHECS--------YPMCKSCFEFEIKEGRKVC 51
              | :||           |  ||      |  ||:|:          |:|  |:: :  | :
Db        718 CKSCGSTTPGKGWDAQWSHD-----FSLCHDCAKLFAKGNFCPLCDKCYDDDDYESK--M 770

Qy         52 LRCG 55
              ::||
Db        771 MQCG 774


RESULT 11
US-08-061-376-5
 Sequence 5, Application US/08061376
 Patent No. 6175000
  GENERAL INFORMATION:
    APPLICANT:  Evans, Glen A.
    APPLICANT:  Djabali, Malek
;    APPLICANT:  Selleri, Licia
;    APPLICANT:  Parry, Pauline
;    TITLE OF INVENTION:  CHARACTERIZATION OF A CHROMOSOME 11Q23
;    TITLE OF INVENTION:  TRANSLOCATION BREAKPOINT ASSOCIATED WITH ACUTE LEUKEMIAS
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pretty, Schroeder, Brueggemann & Clark
;      STREET:  444 South Flower Street, Suite 2000
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  90071
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/061,376
;      FILING DATE:  13-MAY 1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Reiter, Stephen E.
;      REGISTRATION NUMBER:  31,192
;      REFERENCE/DOCKET NUMBER:  P41 9387
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619)546-4737
;      TELEFAX:  (619)546-9392
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3969 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-08-061-376-5

  Query Match             20.7%;  Score 67;  DB 3;  Length 3969;
  Best Local Similarity   25.0%;  Pred. No. 30;
  Matches   16;  Conservative   10;  Mismatches   14;  Indels   24;  Gaps    4;

Qy          9 CHTCGE---------QVGHDANGELFVACHECS--------YPMCKSCFEFEIKEGRKVC 51
              | :||           |  ||      |  ||:|:          |:|  |:: :  | :
Db       1532 CKSCGSTTPGKGWDAQWSHD---  FSLCHDCAKLFAKGNFCPLCDKCYDDDDYESK--M 1584

Qy         52 LRCG 55
              ::||
Db       1585 MQCG 1588


RESULT 12
US-08-678-039A-42
; Sequence 42, Application US/08678039A
; Patent No. 5858662
;  GENERAL INFORMATION:
;    APPLICANT:  Keating, Mark T.
;    APPLICANT:  Morris, Colleen A.
;    TITLE OF INVENTION:  Diagnosis of Williams Syndrome and
;    TITLE OF INVENTION:  Williams Syndrome Cognitive Profile by Analysis of the
;    TITLE OF INVENTION:  Presence or Absence of a LIM-Kinase Gene
;    NUMBER OF SEQUENCES:  42
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Rothwell, Figg, Ernst & Kurz, P.C.
;      STREET:  555 Thirteenth Street, N.W., Suite 701 East
;      STREET:  Tower
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  U.S.A.
;      ZIP:  20004
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
```

```
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/678,039A
;      FILING DATE:  10-JUL-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Saxe, Stephen A.
;      REGISTRATION NUMBER:  38,609
;      REFERENCE/DOCKET NUMBER:  2323-120A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-624-1589
;      TELEFAX:  202-783-6031
;  INFORMATION FOR SEQ ID NO:  42:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  647 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-678-039A-42

  Query Match             20.4%;  Score 66;  DB 2;  Length 647;
  Best Local Similarity   18.9%;  Pred. No. 6;
  Matches   21;  Conservative   14;  Mismatches   12;  Indels   64;  Gaps    5;

Qy          1 MMESGA--PICHTCGEQV----------------------  ------GHDANGELFV-- 26
                |  |  |:  |:|  :||:::                              :: :|:||
Db         15 MGEEGSELPVCASCGQRIYDGQYLQALNADWHADCFRCCDCSASLSHQYYEKDGQLFCKK 74

Qy         27 --------ACHECSYPMCK-------------SCFEFEIKEGRKVCLRCGS 56
                     :|| ||   : |               ||         :|| ||:
Db         75 DYWARYGESCHGCSEQITKGLVMVAGELKYHPECF---------ICLTCGT 116


RESULT 13
US-08-946-026-31
; Sequence 31, Application US/08946026
; Patent No. 6034218
;  GENERAL INFORMATION:
;    APPLICANT:  Reed, Steven G.
;    APPLICANT:  Dillon, Davin C.
;    APPLICANT:  Twardzik, Daniel R.
;    APPLICANT:  Mitcham, Jennifer L.
;    TITLE OF INVENTION:  COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
;    TITLE OF INVENTION:  AND IMMUNODIAGNOSIS OF PROSTATE CANCER
;    NUMBER OF SEQUENCES:  59
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/946,026
;      FILING DATE:  07-OCT-1997
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
;      REGISTRATION NUMBER:  31,392
;      REFERENCE/DOCKET NUMBER:  210121.424C1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  31:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  124 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
US-08-946-026-31

  Query Match             20.2%;  Score 65.5;  DB 3;  Length 124;
  Best Local Similarity   29.3%;  Pred. No. 1.2;
  Matches   17;  Conservative    5;  Mismatches   23;  Indels   13;  Gaps    2;

Qy          6 APICHTCGEQVGHDANGELFVACHECSYPMCKSC---------FEFEIKEGRKVCLRC 54
              || || ||| :     | :  | :   :| |  |              |    || :|  |
Db         64 APCCHQCGEFI----IGRVIKAMNNSWHPECFRCDLCQEVLADIGFVKNAGRHLCRPC 117


RESULT 14
US-08-269-441A-2
; Sequence 2, Application US/08269441A
; Patent No. 5552529
;  GENERAL INFORMATION:
;    APPLICANT:  Rearden, Ann
;    TITLE OF INVENTION:  A NOVEL AUTOANTIGEN, PINCH
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  4225 Executive Square, Suite 1400
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92037
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/269,441A
;      FILING DATE:  30-JUN-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Haile Ph.D., Lisa A.
;      REGISTRATION NUMBER:  38,347
;      REFERENCE/DOCKET NUMBER:  07257/009001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 678-5070
;      TELEFAX:  (619) 678-5099
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  314 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-269-441A-2

  Query Match             20.2%;  Score 65.5;  DB 1;  Length 314;
  Best Local Similarity   29.3%;  Pred. No. 3.2;
  Matches   17;  Conservative    5;  Mismatches   23;  Indels   13;  Gaps    2;

Qy          6 APICHTCGEQVGHDANGELFVACHECSYPMCKSC---------FEFEIKEGRKVCLRC 54
              || || ||| :     | :  | :   :| |  |              |    || :|  |
Db         68 APCCHQCGEFI----IGRVIKAMNNSWHPECFRCDLCQEVLADIGFVKNAGRHLCRPC 121


RESULT 15
US-09-008-465-3
; Sequence 3, Application US/09008465
; Patent No. 6174702
;  GENERAL INFORMATION:
;    APPLICANT:  Lal, Preeti
;    APPLICANT:  Guegler, Karl J.
;    APPLICANT:  Corley, Neil C.
;    TITLE OF INVENTION:  HUMAN PINCH PROTEIN HOMOLOG
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Dr.
```

CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/008,465
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0460 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 314 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 1516012
S-09-008-465-3

Query Match 20.2%; Score 65.5; DB 3; Length 314;
Best Local Similarity 29.3%; Pred. No. 3.2;
Matches 17; Conservative 5; Mismatches 23; Indels 13; Gaps 2;

```
y           6 APICHTCGEQVGHDANGELFVACHECSYPMCKSC---------FEFEIKEGRKVCLRC 54
              || || ||| :     | :  | :   :| |  |           |    || :|  !
b          68 APCCHQCGEFI----IGRVIKAMNNSWHPECFRCDLCQEVLADIGFVKNAGRHLCRPC 121
```

earch completed: October 29, 2003, 18:32:04
ob time : 2.62476 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 29, 2003, 18:24:13 ; Search time 4.38685 Seconds
(without alignments)
2026.213 Million cell updates/sec

Title: US-09-980-043-4
Perfect score: 324
Sequence: 1 MMESGAPICHTCGEQVGHDA.....KSCFEPEIKEGRKVCLRCGS 56

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	324	100.0	56	22	AAB37894	Populus tremuloide
2	264	81.5	974	19	AAW60039	Cotton cellulose s
3	256	79.0	974	19	AAW73308	Cellulose synthase
4	253	78.1	946	22	AAB37893	Populus tremuloide
5	246	75.9	958	23	ABB92981	Herbicidally activ
6	233	71.9	184	21	AAB16320	Eucalyptus grandis
7	233	71.9	251	21	AAB16274	Eucalyptus grandis
8	233	71.9	955	21	AAB16319	Eucalyptus grandis
9	165	50.9	245	21	AAB16308	Pinus radiata cell
10	165	50.9	603	21	AAB16307	Pinus radiata cell
11	165	50.9	1043	23	ABB93701	Herbicidally activ
12	164	50.6	1039	19	AAW73309	Cellulose synthase
13	162	50.0	1075	21	AAY84109	Amino acid sequenc
14	162	50.0	1075	21	AAY84115	AMino acid sequenc
15	162	50.0	1075	21	AAY84120	Amino acid sequenc
16	161	49.7	547	19	AAW46202	Rice RSW1-like cel
17	161	49.7	1074	21	AAY84110	Amino acid sequenc
18	161	49.7	1074	21	AAY84112	Amino acid sequenc
19	161	49.7	1074	21	AAY84117	Amino acid sequenc
20	158	48.8	1058	21	AAY84107	Amino acid sequenc
21	158	48.8	1148	21	AAY58832	Corn cellulose syn
22	156	48.1	354	19	AAW73310	Cellulose synthase
23	155	47.8	1065	23	ABB91880	Herbicidally activ
24	154	47.5	1039	21	AAY58837	Soybean cellulose
25	153.5	47.4	1094	21	AAY84113	Amino acid sequenc
26	153.5	47.4	1094	21	AAY84118	Amino acid sequenc
27	153.5	47.4	1165	21	AAY58834	Corn cellulose syn
28	152	46.9	1026	23	ABB93522	Herbicidally activ
29	152	46.9	1081	19	AAW33817	Arabidopsis cellul
30	152	46.9	1081	19	AAW33820	Arabidopsis cellul
31	152	46.9	1081	23	ABB93199	Herbicidally activ
32	150	46.3	1086	21	AAY84111	Amino acid sequenc
33	150	46.3	1086	21	AAY84116	Amino acid sequenc
34	150	46.3	1086	21	AAY58835	Corn cellulose syn
35	148	45.7	1043	21	AAY84119	Amino acid sequenc
36	148	45.7	1077	21	AAY84108	Amino acid sequenc
37	148	45.7	1077	21	AAY84114	Amino acid sequenc
38	148	45.7	1079	23	AAU10496	Corn cellulose syn
39	148	45.7	1084	19	AAW33818	Arabidopsis cellul
40	148	45.7	1084	23	ABB93304	Herbicidally activ
41	146	45.1	1069	23	ABB93411	Herbicidally activ
42	146	45.1	1076	21	AAY84121	DNA encoding a mai
43	144	44.4	1088	23	ABB91831	Herbicidally activ
44	139	42.9	1080	22	AAB37895	Arabidopsis thalia
45	139	42.9	1084	23	ABB93949	Herbicidally activ

ALIGNMENTS

RESULT 1
AAB37894
ID AAB37894 standard; Protein; 56 AA.
XX
AC AAB37894;
XX
DT 07-MAR-2001 (first entry)
XX
DE Populus tremuloides CelA N-terminal sequence.
XX
KW Aspen; CelA; cellulose synthase; arabidopsis; wood;
KW mechanical stress response element; MSRE; cellulose biosynthesis.
XX
OS Populus tremuloides.
XX
PN WO200071670-A2.
XX
PD 30-NOV-2000.
XX
PF 18-MAY-2000; 2000WO-US13637.
XX
PR 21-MAY-1999; 99US-0135280.
XX
PA (UNMT) UNIV MICHIGAN TECHNOLOGICAL.
XX
PI Chiang VL, Wu L, Chandrashekhar JP, Carraway DT;
XX
DR WPI; 2001-032025/04.
DR N-PSDB; AAC65449.
XX
PT Novel polynucleotide encoding cellulose synthase useful for enhancing
PT cellulose biosynthesis, modifying lignin biosynthesis, and for

```
T   accelerating growth of plants and for upregulating cellulose synthase
T   levels -
X
S   Claim 1; Fig 4; 65pp; English.
X
C   The present sequence is given in a specification relating to
C   polynucleotides encoding cellulose synthase enzymes from Populus
C   tremuloides and Arabidopsis. The cellulose synthases are designated
C   PtCelA and AraxCelA respectively. The polynucleotides may be introduced
C   into plants where they are useful for altering the growth of a plant.
C   Vectors comprising a cellulose synthase promoter, or a fragment of the
C   promoter containing one or more mechanical stress response elements
C   (MSRE), operably linked to a cellulose synthase gene are useful for
C   causing stress-induced gene expression in a plant cell. A cDNA encoding a
C   protein that binds to a positive MSRE may be delivered into the plant to
C   increase expression of cellulose synthase. A cDNA in an antisense
C   orientation may be delivered to reduce expression of cellulose synthase.
C   The polynucleotides enable the development of transgenic tree varieties
C   having increased cellulose content, decreased lignin content and
C   therefore improved wood fibre characteristics. The cellulose synthase
C   promoters are useful for synthesising greater quantities of
C   high-crystalline cellulose in plant and preferably in trees, which
C   permits an increase in proportion of cellulose in transgenic plants,
C   greater strength of juvenile wood and fiber and acceleration of overall
C   growth rate.
C   Note: This polypeptide sequence is given in the specification as
C   SEQ ID NO: 4. SEQ ID NO: 4 is referred to in claim 1 but it is described
C   as a nucleotide sequence.
X
Q   Sequence   56 AA;

 Query Match             100.0%;  Score 324;  DB 22;  Length 56;
 Best Local Similarity   100.0%;  Pred. No. 1.1e-28;
 Matches   56;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MMESGAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEIKEGRKVCLRCGS 56
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MMESGAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEIKEGRKVCLRCGS 56


ESULT 2
AW60039
D   AAW60039 standard; Protein; 974 AA.
X
C   AAW60039;
X
T   28-SEP-1998  (first entry)
X
E   Cotton cellulose synthase CelA1.
X
W   Cellulose synthase; CelA1; cotton; forest tree; plant fibre;
W   wood; transgenic plant; ss.
X
S   Gossypium hirsutum Acala SJ-2.
X
H   Key             Location/Qualifiers
T   Region          175..295
T                   /label= H-1
T   Region          280..295
T                   /label= U-1
T   Region          296..420
T                   /label= P-CR
T   Region          421..539
T                   /label= H-2
T   Region          450..467
T                   /label= U-2
T   Region          540..658
T                   /label= HVR
T   Region          659..880
T                   /label= H-3
T   Region          667..688
T                   /label= U-3
T   Region          706..722
FT                  /label= U-4
FT   Modified-site  81
FT                  /note= "N-glycosylated"
FT   Modified site  466
FT                  /note= "N glycosylated"
FT   Modified site  703
FT                  /note= "N-glycosylated"
FT   Modified-site  846
FT                  /note= "N-glycosylated"
FT   Misc-difference 293
FT                  /note= "critical Asp residue"
FT   Misc-difference 461
FT                  /note= "critical Asp residue"
FT   Misc-difference 672
FT                  /note= "critical Asp residue"
FT   Peptide        710..714
FT                  /note= "QXXRW motif"
XX
PN   WO9818949-A2.
XX
PD   07-MAY-1998.
XX
PF   29-OCT-1997;   97WO-US019629.
XX
PR   29-OCT 1996;   96US-0029987
XX
PA   (CALJ ) CALGENE INC.
XX
PI   Delmer DP,  Pear JR,  Stalker DM;
XX
DR   WPI; 1998-272238/24.
DR   N-PSDB; AAV34432.
XX
PT   New isolated plant cellulose synthesis enzyme DNA - is used to
PT   develop products for producing modified plant fibre products,
PT   particularly for modifying cotton fibre and wood quality
XX
PS   Example 5; Fig 3A-C; 49pp; English.
XX
CC   This full-length polypeptide comprises cotton cellulose synthase
CC   CelA1.  Its amino acid sequence was deduced from a cDNA clone (see
CC   AAV60039) isolated from a cotton fibre cDNA library.  Expression of
CC   CelA1 and CelA2 (see AAW60040) in developing fibre is approximately
CC   500 times higher than the level of expression in other cotton
CC   tissues; approximately 1-2% of the 24 dpa mRNA is CelA mRNA.
CC   Sequence analysis of CelA1 and CelA2 showed regions of high homology
CC   (H-1, H-2 and H-3) between plant and bacteral proteins, as well as
CC   a region (P-CR) that is conserved among plant CelA genes, and
CC   another region (HVR) that is hypervariable between plant genes.
CC   Regions U-1 to U-4 are predicted to be critical for UDP-glc binding
CC   and catalysis in bacterial CelA proteins.  The isolated CelA cDNA
CC   clones are used in claimed methods for modifying fibre phenotype
CC   in a cotton plant and for modifying wood quality phenotype in a
CC   forest tree species.  Lignin can be reduced by over-expressing the
CC   cellulose synthase product and diverting production into cellulose.
XX
SQ   Sequence   974 AA;

  Query Match             81.5%;  Score 264;  DB 19;  Length 974;
  Best Local Similarity   75.0%;  Pred. No. 8.6e-21;
  Matches   42;  Conservative    8;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          1 MMESGAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEIKEGRKVCLRCGS 56
              ||||| |:|||||| || : ||| ||||| |:|:|||||:::|||| ||||||
Db          1 MMESGVPVCHTCGEHVGLNVNGEPFVACHECNFPICKSCFEYDLKEGRKACLRCGS 56


RESULT 3
AAW73308
ID   AAW73308 standard; Protein; 974 AA.
XX
AC   AAW73308;
XX
```

DT 08-FEB-1999 (first entry)
XX
DE Cellulose synthase PcsA1.
XX
KW Cellulose synthase; cotton.
XX
OS Gossypium hirsutum.
XX
PN EP875575-A2.
XX
PD 04-NOV-1998.
XX
PF 31-MAR-1998; 98EP-0302489.
XX
PR 01-APR-1997; 97JP-0083133.
XX
PA (NISN) NISSHINBO IND INC.
XX
PI Aotsuka S, Hasegawa O, Hayashi T, Ihara Y;
XX
DR WPI; 1998-559440/48.
DR N-PSDB; AAV08372.
XX
PT New DNA encoding cotton cellulose synthase proteins - useful for
PT controlling cellulose synthesis in prokaryotic or eukaryotic cells
XX
PS Claim 1; Page 14-18; 51pp; English.
XX
CC This sequence is a cellulose synthase of the invention. The DNA can
CC be used for controlling cellulose synthesis in a cell by introducing the
CC DNA into the cell, expressing RNA with a nucleotide sequence homologous
CC to the DNA or a nucleotide sequence complementary to the DNA. The cells
CC are prokaryotic cells i.e. acetobacterium and/or eukaryotic cells such as
CC yeasts i.e. Saccharomyces spp., plant cells (cotton plants) and cultured
CC cells of mammals etc.
XX
SQ Sequence 974 AA;

Query Match 79.0%; Score 256; DB 19; Length 974;
Best Local Similarity 72.7%; Pred. No. 6.6e-20;
Matches 40; Conservative 9; Mismatches 6; Indels 0; Gaps 0;

Qy 1 MMESGAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEIKEGRKVCLRCG 55
Db 1 MMESGVPVCHTCGEHVGLNVNGEFFVACHECNFPICKSCFEYDLKEGQKACLRCG 55

RESULT 4
AAB37893
ID AAB37893 standard; Protein; 94[illegible] AA.
XX
AC AAB37893;
XX
DT 07-MAR-2001 (first entry)
XX
DE Populus tremuloides cellulose synthase CelA.
XX
KW Aspen; CelA; cellulose synthase; arabidopsis; wood;
KW mechanical stress response element; MSRE; cellulose biosynthesis.
XX
OS Populus tremuloides.
XX
PN WO200071670-A2.
XX
PD 30-NOV-2000.
XX
PF 18-MAY-2000; 2000WO-US13637.
XX
PR 21-MAY-1999; 99US-0135280.
XX
PA (UNMT) UNIV MICHIGAN TECHNOLOGICAL.
XX
PI Chiang VL, Wu L, Chandrashekhar JP, Carraway DT;
XX
DR WPI; 2001-032025/04.
DR N-PSDB; AAC65448.
XX
PT Novel polynucleotide encoding cellulose synthase useful for enhancing
PT cellulose biosynthesis, modifying lignin biosynthesis, and for
PT accelerating growth of plants and for upregulating cellulose synthase
PT levels -
XX
PS Claim 8; Fig 1; 65pp; English.
XX
CC The present sequence is given in a specification relating to
CC polynucleotides encoding cellulose synthase enzymes from Populus
CC tremuloides and Arabidopsis. The cellulose synthases are designated
CC PtCelA and AraxCelA respectively. The polynucleotides may be introduced
CC into plants where they are useful for altering the growth of a plant.
CC Vectors comprising a cellulose synthase promoter, or a fragment of the
CC promoter containing one or more mechanical stress response elements
CC (MSRE), operably linked to a cellulose synthase gene are useful for
CC causing stress-induced gene expression in a plant cell. A cDNA encoding a
CC protein that binds to a positive MSRE may be delivered into the plant to
CC increase expression of cellulose synthase. A cDNA in an antisense
CC orientation may be delivered to reduce expression of cellulose synthase.
CC The polynucleotides enable the development of transgenic tree varieties
CC having increased cellulose content, decreased lignin content and
CC therefore improved wood fibre characteristics. The cellulose synthase
CC promoters are useful for synthesising greater quantities of
CC high-crystalline cellulose in plant and preferably in trees, which
CC permits an increase in proportion of cellulose in transgenic plants,
CC greater strength of juvenile wood and fiber and acceleration of overall
CC growth rate.
XX
SQ Sequence 946 AA;

Query Match 78.1%; Score 253; DB 22; Length 946;
Best Local Similarity 87.5%; Pred. No. 1.4e-19;
Matches 49; Conservative 0; Mismatches 1; Indels 6; Gaps 3;

Qy 1 MMESGAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEIKEGRKVCLRCGS 56
Db 1 MMESGAPICHTCGE--GHDANGELFVACHE--YPMCKSCFEFEINE--KVCLRCGS 50

RESULT 5
ABB92981
ID ABB92981 standard; Protein; 958 AA.
XX
AC ABB92981;
XX
DT 31-MAY-2002 (first entry)
XX
DE Herbicidally active polypeptide SEQ ID NO 2192.
XX
KW Herbicidal; plant; agriculture; herbicide.
XX
OS Arabidopsis thaliana.
XX
PN WO200210210-A2.
XX
PD 07-FEB-2002.
XX
PF 28-AUG-2001; 2001WO-EP09892.
XX
PR 28-AUG-2001; 2001WO-EP09892.
XX
PA (FARB) BAYER AG.
XX
PI Tietjen K, Weidler M;
XX
DR WPI; 2002-269010/31.
XX
PT Identifying plant target proteins for herbicidally active compounds,
PT comprising aligning and comparing nucleic acid or amino acid sequences

```
'T   from plant with nucleic acid or amino acid sequences from non-plant
'T   organisms -
X
S   Claim 5; SEQ ID NO 2192; 261pp + Sequence Listing; English.
X
C   The invention relates to identifying target proteins
'C  (ABB90790-ABB94016) for herbicidally active compounds, comprising
'C  aligning and comparing nucleic acid or amino acid sequences from plant
'C  with nucleic acid or amino acid sequences from non-plant organisms using
'C  suitable search parameters, where plant sequences having an E-value
'C  greater by a factor of 3 than the E-value of most similar non-plant
C   sequences are selected. The polypeptides or nucleic acids encoding them
C   are useful for identifying modulators. The identified modulators are
C   useful as herbicides.
X
Q   Sequence   958 AA;

 Query Match             75.9%;  Score 246;  DB 23;  Length 958;
 Best Local Similarity   67.9%;  Pred. No. 8.3e-19;
 Matches   38;  Conservative   12;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          1 MMESGAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEIKEGRKVCLRCGS 56
              |||| :|||:|||||::|  :||| ||||||||:|:||:| |:| ||||::|||||:
Db          1 MMESRSPICNTCGEEIGVKSNGEFFVACHECSFPICKACLEYEFKEGRRICLRCGN 56


ESULT 6
AB16320
D   AAB16320 standard; Protein; 184 AA.
X
C   AAB16320;
X
T   31-OCT-2000  (first entry)
X
E   Eucalyptus grandis cellulose synthase protein sequence SEQ ID NO:108.
X
W   Eucalyptus grandis; pinus radiata; Monterey pine; modification;
W   plant cell wall; polysaccharide; polysaccharide biosynthetic pathway;
W   transgenic plant.
X
S   Eucalyptus grandis.
X
N   WO200022092-A2.
X
D   20-APR-2000.
X
F   08-OCT-1999;   99WO-NZ00169.
X
R   13-OCT-1998;   98US-0170862.
R   11-AUG-1999;   99US-0148426.
X
A   (GENE-) GENESIS RES & DEV CORP LTD.
A   (FLET-) FLETCHER CHALLENGE FORESTS LTD.
X
I   Bloksberg LN;
X
R   WPI; 2000-339328/29.
R   N-PSDB; AAA67127.
X
T   New genes encoding proteins involved in a plant polysaccharide
'T  biosynthetic pathway, useful for modulating or altering the
'T  polysaccharide content, composition or structure of the plant  -
X
S   Claim 17; Page 92-93; 301pp; English.
X
C   The present invention describes isolated polynucleotides (PN) comprising
C   a sequence selected from one of 835 nucleotide sequences given in
C   AAA67073 to AAA67907, their (reverse) complements, sequences producing
C   an Expectation (E) value of 0.01 or less compared to the 835 sequences,
C   sequences at least 50% identical to them, 200, 100, 40 or 20-mers of the
C   835 sequences or sequences that are degenerately equivalent or allelic
C   to the 835 sequences. The polynucleotides are used to modify the
C   activity of a polypeptide involved in a polysaccharide biosynthetic
CC  pathway in the plant. They are especially used to modulate or alter the
CC  polysaccharide content, composition or structure of the plant. AAB16268
CC  to AAB16340 are proteins encoded by some of the polynucleotide sequence
CC  given in the present invention.
XX
SQ  Sequence   184 AA;

  Query Match             71.9%;  Score 233;  DB 21;  Length 184;
  Best Local Similarity   67.9%;  Pred. No. 4.2e-18;
  Matches   38;  Conservative    9;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          1 MMESGAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEIKEGRKVCLRCGS 56
              ||||| |:|:|||| || |  :|:|||| ||:: :||:| |:|||||| ||||||:
Db          1 MMESGVPLCNTCGEAVGVDEKGEVFVACQECNFAICKACVEYEIKEGRKACLRCGT 56


RESULT 7
AAB16274
ID  AAB16274 standard; Protein; 251 AA.
XX
AC  AAB16274;
XX
DT  31-OCT-2000  (first entry)
XX
DE  Eucalyptus grandis cellulose synthase protein sequence SEQ ID NO:36.
XX
KW  Eucalyptus grandis; pinus radiata; Monterey pine; modification;
KW  plant cell wall; polysaccharide; polysaccharide biosynthetic pathway;
KW  transgenic plant.
XX
OS  Eucalyptus grandis.
XX
PN  WO200022092-A2.
XX
PD  20-APR-2000.
XX
PF  08-OCT-1999;   99WO-NZ00169.
XX
PR  13-OCT-1998;   98US-0170862.
PR  11-AUG-1999;   99US-0148426.
XX
PA  (GENE-) GENESIS RES & DEV CORP LTD.
PA  (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI  Bloksberg LN;
XX
DR  WPI; 2000-339328/29.
DR  N-PSDB; AAA67079.
XX
PT  New genes encoding proteins involved in a plant polysaccharide
PT  biosynthetic pathway, useful for modulating or altering the
PT  polysaccharide content, composition or structure of the plant  -
XX
PS  Claim 17; Page 52-53; 301pp; English.
XX
CC  The present invention describes isolated polynucleotides (PN) comprising
CC  a sequence selected from one of 835 nucleotide sequences given in
CC  AAA67073 to AAA67907, their (reverse) complements, sequences producing
CC  an Expectation (E) value of 0.01 or less compared to the 835 sequences,
CC  sequences at least 50% identical to them, 200, 100, 40 or 20-mers of the
CC  835 sequences or sequences that are degenerately equivalent or allelic
CC  to the 835 sequences. The polynucleotides are used to modify the
CC  activity of a polypeptide involved in a polysaccharide biosynthetic
CC  pathway in the plant. They are especially used to modulate or alter the
CC  polysaccharide content, composition or structure of the plant. AAB16268
CC  to AAB16340 are proteins encoded by some of the polynucleotide sequence
CC  given in the present invention.
XX
SQ  Sequence   251 AA;

  Query Match             71.9%;  Score 233;  DB 21;  Length 251;
  Best Local Similarity   67.9%;  Pred. No. 5.8e-18;
  Matches   38;  Conservative    9;  Mismatches    9;  Indels    0;  Gaps    0;
```

```
Qy          1 MMESGAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEIKEGRKVCLRCGS 56
              ||||| |:|:|||| || |  :|:|||| ||:: :||:| |:|||||| |||||:
Db          1 MMESGVPLCNTCGEAVGVDEKGEVFVACQECNFAICKACVEYEIKEGRKACLRCGT 56


RESULT 8
AAB16338
ID   AAB16338 standard; Protein; 955 AA.
XX
AC   AAB16338;
XX
DT   31-OCT-2000  (first entry)
XX
DE   Eucalyptus grandis cellulose synthase protein sequence SEQ ID NO:146.
XX
KW   Eucalyptus grandis; pinus radiata; Monterey pine; modification;
KW   plant cell wall; polysaccharide; polysaccharide biosynthetic pathway;
KW   transgenic plant.
XX
OS   Eucalyptus grandis.
XX
PN   WO200022092-A2.
XX
PD   20-APR-2000.
XX
PF   08-OCT-1999;   99WO-NZ00169.
XX
PR   13-OCT-1998;   98US-0170862.
PR   11-AUG-1999;   99US-0148426.
XX
PA   (GENE-) GENESIS RES & DEV CORP LTD.
PA   (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI   Bloksberg LN;
XX
DR   WPI; 2000-339328/29.
DR   N-PSDB; AAA67145.
XX
PT   New genes encoding proteins involved in a plant polysaccharide
PT   biosynthetic pathway, useful for modulating or altering the
PT   polysaccharide content, composition or structure of the plant  -
XX
PS   Claim 17; Page 111-113; 301pp; English
XX
CC   The present invention describes isolated polynucleotides (PN) comprising
CC   a sequence selected from one of 835 nucleotide sequences given in
CC   AAA67073 to AAA67907, their (reverse) complements, sequences producing
CC   an Expectation (E) value of 0.01 or less compared to the 835 sequences,
CC   sequences at least 50% identical to them, 200, 100, 40 or 20-mers of the
CC   835 sequences or sequences that are degenerately equivalent or allelic
CC   to the 835 sequences. The polynucleotides are used to modify the
CC   activity of a polypeptide involved in a polysaccharide biosynthetic
CC   pathway in the plant. They are especially used to modulate or alter the
CC   polysaccharide content, composition or structure of the plant. AAB16268
CC   to AAB16340 are proteins encoded by some of the polynucleotide sequence
CC   given in the present invention.
XX
SQ   Sequence   955 AA;

  Query Match             71.9%;  Score 223;  DB 21;  Length 955;
  Best Local Similarity   67.9%;  Pred. No. 2.2e-17;
  Matches   38;  Conservative    9;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          1 MMESGAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEIKEGRKVCLRCGS 56
              |||'| |:|::||| |' |  : .|| | ||:: :||:| |:||:|||| |||||:
Db          1 MMESGVPLCNTCGEAVGVDEKGEVFVACQECNFAICKACVEYEIKEGRKACLRCGT 56


RESULT 9
AAB16308
ID   AAB16308 standard; Protein; 245 AA.
XX
AC   AAB16308;
XX
DT   31-OCT-2000  (first entry)
XX
DE   Pinus radiata cellulose synthase protein sequence SEQ ID NO:94.
XX
KW   Eucalyptus grandis; pinus radiata; Monterey pine; modification;
KW   plant cell wall; polysaccharide; polysaccharide biosynthetic pathway;
KW   transgenic plant.
XX
OS   Pinus radiata.
XX
PN   WO200022092-A2.
XX
PD   20-APR-2000.
XX
PF   08-OCT-1999;   99WO-NZ00169.
XX
PR   13-OCT-1998;   98US-0170862.
PR   11-AUG-1999;   99US-0148426.
XX
PA   (GENE-) GENESIS RES & DEV CORP LTD.
PA   (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI   Bloksberg LN;
XX
DR   WPI; 2000-339328/29.
DR   N-PSDB; AAA67115.
XX
PT   New genes encoding proteins involved in a plant polysaccharide
PT   biosynthetic pathway, useful for modulating or altering the
PT   polysaccharide content, composition or structure of the plant  -
XX
PS   Claim 17; Page 84-85; 301pp; English.
XX
CC   The present invention describes isolated polynucleotides (PN) comprising
CC   a sequence selected from one of 835 nucleotide sequences given in
CC   AAA67073 to AAA67907, their (reverse) complements, sequences producing
CC   an Expectation (E) value of 0.01 or less compared to the 835 sequences,
CC   sequences at least 50% identical to them, 200, 100, 40 or 20-mers of the
CC   835 sequences or sequences that are degenerately equivalent or allelic
CC   to the 835 sequences. The polynucleotides are used to modify the
CC   activity of a polypeptide involved in a polysaccharide biosynthetic
CC   pathway in the plant. They are especially used to modulate or alter the
CC   polysaccharide content, composition or structure of the plant. AAB16268
CC   to AAB16340 are proteins encoded by some of the polynucleotide sequence
CC   given in the present invention.
XX
SQ   Sequence   245 AA;

  Query Match             50.9%;  Score 165;  DB 21;  Length 245;
  Best Local Similarity   50.0%;  Pred. No. 1.8e-10;
  Matches   23;  Conservative   13;  Mismatches   10;  Indels    0;  Gaps    0;

Qy          9 CHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEIKEGRKVCLRC 54
              |  ||:::| : ||||||||:|| :|:|: |:|:| |:| : | :|
Db         53 CQVCGDKIGCNPNGELFVACNECGFPVCRPCYEYERKDGNRCCPQC 98


RESULT 10
AAB16307
ID   AAB16307 standard; Protein; 603 AA.
XX
AC   AAB16307;
XX
DT   31-OCT-2000  (first entry)
XX
DE   Pinus radiata cellulose synthase protein sequence SEQ ID NO:93.
XX
KW   Eucalyptus grandis; pinus radiata; Monterey pine; modification;
KW   plant cell wall; polysaccharide; polysaccharide biosynthetic pathway;
KW   transgenic plant.
XX
```

Pinus radiata.

WO200022092-A2.

20-APR-2000.

08-OCT-1999; 99WO-NZ00169.

13-OCT-1998; 98US-0170862.
11-AUG-1999; 99US-0148426.

(GENE-) GENESIS RES & DEV CORP LTD.
(FLET-) FLETCHER CHALLENGE FORESTS LTD.

Bloksberg LN;

WPI; 2000-339328/29.
N-PSDB; AAA67114.

New genes encoding proteins involved in a plant polysaccharide biosynthetic pathway, useful for modulating or altering the polysaccharide content, composition or structure of the plant

Claim 17; Page 83-84; 301pp; English.

The present invention describes isolated polynucleotides (PN) comprising a sequence selected from one of 835 nucleotide sequences given in AAA67073 to AAA67907, their (reverse) complements, sequences producing an Expectation (E) value of 0.01 or less compared to the 835 sequences, sequences at least 50% identical to them, 200, 100, 40 or 20-mers of the 835 sequences or sequences that are degenerately equivalent or allelic to the 835 sequences. The polynucleotides are used to modify the activity of a polypeptide involved in a polysaccharide biosynthetic pathway in the plant. They are especially used to modulate or alter the polysaccharide content, composition or structure of the plant. AAB16268 to AAB16340 are proteins encoded by some of the polynucleotide sequence given in the present invention.

Sequence 603 AA;

```
Query Match             50.9%;  Score 165;  DB 21;  Length 603;
Best Local Similarity   49.0%;  Pred. No. 4.6e-10;
Matches   24;  Conservative   13;  Mismatches   12;  Indels    0;  Gaps    0;

          8 ICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEIKEGRKVCLRCGS 56
            :|  ||| || : :|||||||:|| :|:|: |:|:| :|| : | :| :
         49 VCQICGEDVGLNTDGELFVACNECGFPVCRPCYEYERREGNQSCPQCNT 97
```

SULT 11
B93701
ABB93701 standard; Protein; 1043 AA.

ABB93701;

31-MAY-2002 (first entry)

Herbicidally active polypeptide SEQ ID NO 2912.

Herbicidal; plant; agriculture; herbicide.

Arabidopsis thaliana.

WO200210210-A2.

07-FEB-2002.

28-AUG-2001; 2001WO-EP09892.

28-AUG-2001; 2001WO-EP09892.

(FARB) BAYER AG.

PI Tietjen K, Weidler M;
XX
DR WPI; 2002-269310/31.
XX
PT Identifying plant target proteins for herbicidally active compounds, comprising aligning and comparing nucleic acid or amino acid sequences from plant with nucleic acid or amino acid sequences from non-plant organisms -
XX
PS Claim 5; SEQ ID NO 2912; 261pp + Sequence Listing; English.
XX
CC The invention relates to identifying target proteins (ABB90790-ABB94016) for herbicidally active compounds, comprising aligning and comparing nucleic acid or amino acid sequences from plant with nucleic acid or amino acid sequences from non-plant organisms using suitable search parameters, where plant sequences having an E-value greater by a factor of > than the E-value of most similar non-plant sequences are selected. The polypeptides or nucleic acids encoding them are useful for identifying modulators. The identified modulators are useful as herbicides
XX
SQ Sequence 1043 AA;

```
  Query Match             50.9%;  Score 165;  DB 23;  Length 1043;
  Best Local Similarity   51.0%;  Pred. No. 8e-10;
  Matches   26;  Conservative    8;  Mismatches   17;  Indels    0;  Gaps    0;

Qy          6 APICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEIKEGRKVCLRCGS 56
              | ||  ||:|  | |:: |:|| | .|:|. :|:|   | | | :| :
Db         15 AKICKVCGDEVKDDDNGQTFVACHVCVYPVCRPCYEYERSNGNKCCPQCNT 65
```

RESULT 12
AAW73309
ID AAW73309 standard; Protein; 1039 AA
XX
AC AAW73309;
XX
DT 08-FEB-1999 (first entry)
XX
DE Cellulose synthase PesA2.
XX
KW Cellulose synthase; cotton.
XX
OS Gossypium hirsutum.
XX
PN EP875575-A2.
XX
PD 04-NOV-1998.
XX
PF 31-MAR-1998; 98EP-0302489.
XX
PR 01-APR-1997; 97JP-0083133.
XX
PA (NISN) NISSHINBO IND INC.
XX
PI Aotsuka S, Hasegawa O, Hayashi T, Ihara Y;
XX
DR WPI; 1998-559440/48.
DR N-PSDB; AAV08373.
XX
PT New DNA encoding cotton cellulose synthase proteins - useful for controlling cellulose synthesis in prokaryotic or eukaryotic cells
XX
PS Claim 1; Page 22-27; 51pp; English.
XX
CC This sequence is a cellulose synthase of the invention. The DNA can
CC be used for controlling cellulose synthesis in a cell by introducing the
CC DNA into the cell, expressing RNA with a nucleotide sequence homologous
CC to the DNA or a nucleotide sequence complementary to the DNA. The cells
CC are prokaryotic cells i.e. acetobacterium and/or eukaryotic cells such as
CC yeasts i.e. Saccharomyces spp., plant cells (cotton plants) and cultured
CC cells of mammals etc.

```
XX
SQ   Sequence   1039 AA;

  Query Match             50.6%;  Score 164;  DB 19;  Length 1039;
  Best Local Similarity   42.6%;  Pred. No. 1e-09;
  Matches   23;  Conservative   14;  Mismatches   17;  Indels    0;  Gaps    0;

Qy          3 ESGAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEIKEGRKVCLRCGS 56
              :|   ||  ||:::|   ||: !|||| |::|:|: |:|:|  || : | :| :
Db         25 QSSTKICRVCGDKIGQKENGQPFVACHVCAFPVCRPCYEYERSEGNQCCPQCNT 78


RESULT 13
AAY84109
ID   AAY84109 standard; Protein; 1075 AA.
XX
AC   AAY84109;
XX
DT   03-JUL-2000  (first entry)
XX
DE   Amino acid sequence of a maize cellulose synthase.
XX
KW   Maize; cellulose synthase; stalk quality; stand; silage; cellulose;
KW   transgenic plant; plant breeding marker.
XX
OS   Zea mays.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 245
FT                   /note= "encoded by NAT"
XX
PN   WO200009706-A2.
XX
PD   24-FEB-2000.
XX
PF   16 AUG-1999;   99WO-US18760.
XX
PR   17-AUG-1998;   98US-0096822.
XX
PA   (PION-) PIONEER HI-BRED INT INC.
XX
PI   Dhugga KS,  Helentjaris TG,  Bowen BA,  Wang X;
XX
DR   WPI; 2000-224343/19.
DR   N-PSDB; AAZ99497.
XX
PT   New genes which encode maize cellulose synthase polypeptides in plants
PT   useful for modulating the expression of cellulose synthase in plants
PT   and to produce transgenic plants expressing the novel protein -
XX
PS   Claim 15; Page 102-105; 119pp; English.
XX
CC   The present sequence represents a maize cellulose synthase polypeptide.
CC   The cellulose synthase can be used for the improvement of stalk quality
CC   for improved stand or silage. It also provides an increased concentration
CC   of cellulose in the pericarp, hardening the kernel and improving its
CC   handling ability. The sequences are used to produce transgenic plants
CC   and seeds expressing the cellulose synthase. The polynucleotide is
CC   used for modulating, preferably increasing, the level of the synthase
CC   in a plant cell. The plants are preferably monocots. The polynucleotide
CC   is also used as a probe or primer in the detection quantitation or
CC   isolation of gene transcripts. The probes are useful in detecting
CC   deficiencies in the level of mRNA in screenings for desired transgenic
CC   plant, for detecting mutations in the gene, for monitoring upregulation
CC   of expression or changes in enzyme activity in screening assays of
CC   compounds, for detection of any number of allelic variants of the gene,
CC   or for use as molecular markers in plant breeding programs. The
CC   isolated nucleic acids of the present invention can also be used for
CC   recombinant expression of their encoded polypeptides or for use as
CC   immunogens in the preparation and/or screening of antibodies. The
CC   proteins can be employed in assays for enzyme agonists or antagonists
CC   of enzyme function or for use of immunogens or antigens to obtain
CC   antibodies specifically immunoreactive with a protein.

XX
SQ   Sequence   1075 AA;

  Query Match             50.0%;  Score 162;  DB 21;  Length 1075;
  Best Local Similarity   48.9%;  Pred. No. 1.8e-09;
  Matches   23;  Conservative   13;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          8 ICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEIKEGRKVCLRC 54
              :|  ||: ||  | |::||||:||::|:|: |:|:| ||| : | :|
Db         40 VCQICGDSVGVSATGDVFVACNECAFPVCRPCYEYERKEGNQCCPQC 86


RESULT 14
AAY84115
ID   AAY84115 standard; Protein; 1075 AA.
XX
AC   AAY84115;
XX
DT   03-JUL-2000  (first entry)
XX
DE   AMino acid sequence of a maize cellulose synthase.
XX
KW   Maize; cellulose synthase; stalk quality; stand; silage; cellulose;
KW   transgenic plant; plant breeding marker.
XX
OS   Zea mays.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 245
FT                   /note= "encoded by NAT"
XX
PN   WO200009706-A2.
XX
PD   24-FEB-2000.
XX
PF   16-AUG-1999;   99WO-US18760.
XX
PR   17-AUG-1998;   98US-0096822.
XX
PA   (PION-) PIONEER HI-BRED INT INC.
XX
PI   Dhugga KS,  Helentjaris TG,  Bowen BA,  Wang X;
XX
DR   WPI; 2000-224343/19.
DR   N-PSDB; AAZ99515.
XX
PT   New genes which encode maize cellulose synthase polypeptides in plants
PT   useful for modulating the expression of cellulose synthase in plants
PT   and to produce transgenic plants expressing the novel protein -
XX
PS   Claim 15; Page 149-152; 119pp; English.
XX
CC   The present sequence represents a maize cellulose synthase polypeptide.
CC   The cellulose synthase can be used for the improvement of stalk quality
CC   for improved stand or silage. It also provides an increased concentration
CC   of cellulose in the pericarp, hardening the kernel and improving its
CC   handling ability. The sequences are used to produce transgenic plants
CC   and seeds expressing the cellulose synthase. The polynucleotide is
CC   used for modulating, preferably increasing, the level of the synthase
CC   in a plant cell. The plants are preferably monocots. The polynucleotide
CC   is also used as a probe or primer in the detection quantitation or
CC   isolation of gene transcripts. The probes are useful in detecting
CC   deficiencies in the level of mRNA in screenings for desired transgenic
CC   plant, for detecting mutations in the gene, for monitoring upregulation
CC   of expression or changes in enzyme activity in screening assays of
CC   compounds, for detection of any number of allelic variants of the gene,
CC   or for use as molecular markers in plant breeding programs. The
CC   isolated nucleic acids of the present invention can also be used for
CC   recombinant expression of their encoded polypeptides or for use as
CC   immunogens in the preparation and/or screening of antibodies. The
CC   proteins can be employed in assays for enzyme agonists or antagonists
CC   of enzyme function or for use of immunogens or antigens to obtain
CC   antibodies specifically immunoreactive with a protein.
```

```
(
Q  Sequence   1075 AA;

 Query Match            50.0%;  Score 162;  DB 21;  Length 1075;
 Best Local Similarity  48.9%;  Pred. No. 1.8e-09;
 Matches   23;  Conservative   13;  Mismatches   11;  Indels    0;  Gaps    0;

y          8 ICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEIKEGRKVCLRC 54
             :|  ||: ||  | |::||||:||::|:|: |:|:| ||| : | :|
D         40 VCQICGDSVGVSATGDVFVACNECAFPVCRPCYEYERKEGNQCCPQC 86


ESULT 15
AY84120
D  AAY84120 standard; Protein; 1075 AA.
(
'  AAY84120;
(
'  03-JUL-2000  (first entry)

C  Amino acid sequence of a maize cellulose synthase.

(
W  Maize; cellulose synthase; stalk quality; stand; silage; cellulose;
W  transgenic plant; plant breeding marker.
(
S  Zea mays.
(
H  Key             Location/Qualifiers
T  Misc-difference 245
T                  /note= "encoded by NAT"
(
W  WO200009706-A2.
(
D  24-FEB-2000.
(
F  16-AUG-1999;   99WO-US18760.
(
R  17-AUG-1998;   98US-0096822.
(
A  (PION-) PIONEER HI-BRED INT INC.
(
I  Dhugga KS,  Helentjaris TG,  Bowen BA,  Wang X;
(
R  WPI; 2000-224343/19.
R  N-PSDB; AAZ99530.
(
T  New genes which encode maize cellulose synthase polypeptides in plants
T  useful for modulating the expression of cellulose synthase in plants
T  and to produce transgenic plants expressing the novel protein  -
(
S  Claim 15; Page 188-191; 119pp; English.
(
C  The present sequence represents a maize cellulose synthase polypeptide.
C  The cellulose synthase can be used for the improvement of stalk quality
C  for improved stand or silage. It also provides an increased concentration
C  of cellulose in the pericarp, hardening the kernel and improving its
C  handling ability. The sequences are used to produce transgenic plants
C  and seeds expressing the cellulose synthase. The polynucleotide is
C  used for modulating, preferably increasing, the level of the synthase
C  in a plant cell. The plants are preferably monocots. The polynucleotide
C  is also used as a probe or primer in the detection quantitation or
C  isolation of gene transcripts. The probes are useful in detecting
C  deficiencies in the level of mRNA in screenings for desired transgenic
C  plant, for detecting mutations in the gene, for monitoring upregulation
C  of expression or changes in enzyme activity in screening assays of
C  compounds, for detection of any number of allelic variants of the gene,
C  or for use as molecular markers in plant breeding programs. The
C  isolated nucleic acids of the present invention can also be used for
C  recombinant expression of their encoded polypeptides or for use as
C  immunogens in the preparation and/or screening of antibodies. The
C  proteins can be employed in assays for enzyme agonists or antagonists
C  of enzyme function or for use of immunogens or antigens to obtain
C  antibodies specifically immunoreactive with a protein.

XX
SQ  Sequence   1075 AA;

  Query Match            50.0%;  Score 162;  DB 21;  Length 1075;
  Best Local Similarity  48.9%;  Pred. No. 1.8e-09;
  Matches   23;  Conservative   13;  Mismatches   11;  Indels    0;  Gaps    0;

Qy         8 ICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEIKEGRKVCLRC 54
             :|  ||: ||  | |::||||:||::|:|: |:|:| ||| : | :|
Db        40 VCQICGDSVGVSATGDVFVACNECAFPVCRPCYEYERKEGNQCCPQC 86


Search completed: October 29, 2003, 18:29:00
Job time : 4.38685 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 31, 2003, 23:20:30 ; Search time 332.263 Seconds
(without alignments)
4096.305 Million cell updates/sec

Title: US-09-980-043-4
Perfect score: 56
Sequence: 1 MMESGAPICHTCGEQVGHDA.....KSCFEFEIKEGRKVCLRCGS 56

Scoring table: OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Word size: 1

Total number of hits satisfying chosen parameters: 45562604

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09980043/runat_29102003_163930_22677/app_query.fasta_1.1358
-DB=EST -QFMT=fastap -SUFFIX=oligo.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09980043_@CGN_1_1_4125_@runat_29102003_163930_22677 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	25	44.6	434	13	BU894130	BU894130 X003C01 P
c 2	9	16.1	512	9	AL368412	AL368412 MtBA24C11
3	9	16.1	515	10	BE326104	BE326104 NF086F04S
4	9	16.1	558	14	CA919485	CA919485 EST637203
5	8	14.3	290	10	BB334832	BB334832 BB334832
6	8	14.3	318	13	BQ701242	BQ701242 NXSI_061_
c 7	8	14.3	385	14	CB702847	CB702847 AMGNNUC:N
c 8	8	14.3	390	14	CB778315	CB778315 AMGNNUC:N
c 9	8	14.3	396	28	AQ610092	AQ610092 HS_5092_A
c 10	8	14.3	403	14	CB768576	CB768576 AMGNNUC:C
c 11	8	14.3	406	14	CB807973	CB807973 AMGNNUC:M
c 12	8	14.3	416	28	AQ443329	AQ443329 HS_2231_B
13	8	14.3	437	13	BY574406	BY574406 BY574406
c 14	8	14.3	470	14	CB001824	CB001824 VVB009E02
c 15	8	14.3	500	28	AQ661271	AQ661271 Sheared D
16	8	14.3	533	28	AZ214004	AZ214004 Sheared D
17	8	14.3	534	10	BE203599	BE203599 EST396275
c 18	8	14.3	541	9	AV399797	AV399797 AV399797
c 19	8	14.3	541	12	BJ053621	BJ053621 BJ053621
c 20	8	14.3	547	12	BJ048677	BJ048677 BJ048677
21	8	14.3	563	12	BM892555	BM892555 sam61f07.
22	8	14.3	582	9	AV830247	AV830247 AV830247
23	8	14.3	585	12	BM893295	BM893295 sam54c08.
c 24	8	14.3	609	14	CB547818	CB547818 AMGNNUC:C
c 25	8	14.3	629	14	CB003702	CB003702 VVB032B08
c 26	8	14.3	631	14	CB916067	CB916067 VVD106C03
c 27	8	14.3	636	14	CB002017	CB002017 VVB012A05
28	8	14.3	644	10	BG449805	BG449805 NF053E04I
c 29	8	14.3	646	14	CB004290	CB004290 VVB037H06
c 30	8	14.3	670	28	AZ215625	AZ215625 Sheared D
31	8	14.3	683	13	BX255611	BX255611 BX255611
c 32	8	14.3	696	14	CB003091	CB003091 VVB024C10
33	8	14.3	700	9	AW299036	AW299036 EST305710
34	8	14.3	701	29	BZ215910	BZ215910 CH230-332
c 35	8	14.3	702	29	BZ698843	BZ698843 PUCDU41TD
36	8	14.3	721	14	CB982729	CB982729 CAB70007_
37	8	14.3	771	14	CB983779	CB983779 AGENCOURT
38	8	14.3	836	29	BZ670304	BZ670304 PUBCC78TD
39	8	14.3	926	28	AZ535033	AZ535033 ENTCU45TF
40	8	14.3	1022	12	BQ063335	BQ063335 AGENCOURT
c 41	7	12.5	134	10	BF836195	BF836195 QV3-HT101
42	7	12.5	134	28	AZ107095	AZ107095 RPCI-23-2
43	7	12.5	149	10	BF174179	BF174179 MYE2706 M
c 44	7	12.5	210	28	B15238	B15238 344G23.TV C
c 45	7	12.5	230	10	AW995831	AW995831 QV4-BN004

ALIGNMENTS

RESULT 1
BU894130
LOCUS BU894130 434 bp mRNA linear EST 17-OCT-2002
DEFINITION X003C01 Populus wood cDNA library Populus tremula x Populus tremuloides cDNA 5 prime, mRNA sequence.
ACCESSION BU894130
VERSION BU894130.1 GI:24105195
KEYWORDS EST.
SOURCE Populus tremula x Populus tremuloides
ORGANISM Populus tremula x Populus tremuloides
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids ; eurosids I; Malpighiales; Salicaceae; Populus.

```
FERENCE   1  (bases 1 to 434)
AUTHORS   Unneberg,P., Bhalerao,R.R., Jansson,S. and Sterky,F.
TITLE     The poplar tree transcriptome: Analysis of expressed sequence tags
          from multiple libraries
JOURNAL   Unpublished
OMMENT    Contact: BHALERAO RUPALI R.
          Umea Plant Science Center
          Department of Plant Physiology
          University of Umea, 901 87 Umea, Sweden
          Tel: +46 90 786 5279
          Fax: +46 90 786 6676
          Email: rupali.bhalerao@plantphys.umu.se.
CATURES             Location/Qualifiers
    source          1. .434
                    /organism="Populus tremula x Populus tremuloides"
                    /mol_type="mRNA"
                    /db_xref="taxon:47664"
                    /tissue_type="wood"
                    /clone_lib="Populus wood cDNA library"
ASE COUNT      132 a     66 c    125 g    111 t
IGIN

ignment Scores:
red. No.:              2.55e-17       Length:        434
ore:                   25.00          Matches:       25
ercent Similarity:     100.00%        Conservative:  0
est Local Similarity:  100.00%        Mismatches:    0
uery Match:            44.64%         Indels:        0
B:                     13             Gaps:          0

S-09-980-043-4 (1-56) x BU894130 (1-434)

Y          1 MetMetGluSerGlyAlaProIleCysHisThrCysGlyGluGlnValGlyHisAspAla 20
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
D         62 ATGATGGAATCTGGGGCTCCTATATGCCATACCTGTGGTGAACAGGTGGGGCATGATGCA 121

Y          21 AsnGlyGluLeuPhe 25
              |||||||||||||||
D         122 AATGGGGAGCTATTT 136


ESULT 2
L368412/c
OCUS        AL368412                 512 bp    mRNA    linear   EST 03-AUG-2000
EFINITION   MtBA24C11R1 MtBA Medicago truncatula cDNA clone MtBA24C11 T7, mRNA
            sequence.
CCESSION    AL368412
ERSION      AL368412.1  GI:9668165
EYWORDS     EST.
OURCE       Medicago truncatula (barrel medic)
 ORGANISM   Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
EFERENCE    1  (bases 1 to 512)
 AUTHORS    Journet,E.P., Crespeau,H., van-Tuinen,D., Gouzy,J., Jaillon,O.,
            Niebel,A., Carreau,V., Chatagnier,O., Kahn,D., Gianinazzi-Pearson
            ,V. and Gamas,P.
 TITLE      Medicago truncatula ESTs from nitrogen-starved roots
 JOURNAL    Unpublished
OMMENT      Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Contact : Pascal Gamas and Etienne-Pascal Journet, Laboratoire de
            Biologie Moleculaire des Relations Plantes-Microorganismes,
            CNRS-INRA, BP 27 31326 Castanet-Tolosan Cedex, France (Email :
            Mt-est@toulouse.inra.fr Website :
            http://sequence.toulouse.inra.fr/Mtruncatula.html).
EATURES             Location/Qualifiers
    source          1. .512
                    /organism="Medicago truncatula"
                    /mol_type="mRNA"
                    /cultivar="Jemalong"
                    /db_xref="taxon:3880"
                    /clone="MtBA24C11"
                    /tissue_type="root tips"
                    /dev_stage="harvested after 3 days of N-starvation"
                    /clone_lib="MtBA"
                    /note="Vector: pBluescript pSK; Site_1: EcoRI; Site_2:
                    XhoI; Plants were grown in an aeroponic chamber for 14
                    days on nitrogen-rich medium followed by 3 days on N-free
                    medium. RNA was extracted from root tips (1-3 cm). cDNA
                    was prepared from polyA+ enriched RNA. The cDNA was
                    directionally ligated into Uni-zapXR vector from
                    Stratagene and packaged using Gigapack Gold packaging
                    extracts. Plasmids containing cDNA inserts were
                    mass-excised from phage stocks using ExAssit helper phage
                    and propagated in SOLR cells. Clone ordering and
                    sequencing was performed by the Centre National de
                    Sequencage (Genoscope, Evry, France)."
BASE COUNT      144 a    126 c     91 g    151 t
ORIGIN

Alignment Scores:
Pred. No.:                 [illegible]     Length:        512
Score:                     9.00            Matches:       9
Percent Similarity:        100.00%         Conservative:  0
Best Local Similarity:     100.00%         Mismatches:    0
Query Match:               16.07%          Indels:        0
DB:                        9               Gaps:          0

US-09-980-043-4 (1-56) x AL368412 (1-512)

Qy          21 AsnGlyGluLeuPheValAlaCysHis 29
               |||||||||||||||||||||||||||
Db         356 AATGGAGAGCTTTTTGTGGCTTGTCAT 330

RESULT 3
BE326104
LOCUS       BE326104                 515 bp    mRNA    linear   EST 21-DEC-2000
DEFINITION  NF086F04ST1F1032 Developing stem Medicago truncatula cDNA clone
            NF086F04ST 5', mRNA sequence.
ACCESSION   BE326104
VERSION     BE326104.2  GI:11958873
KEYWORDS    EST.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 515)
  AUTHORS   He,X.-Z., Shadle,G., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell
            ,C.J., Flores,H.R., Inman,J.T., Weller,J.W., May,G.D. and Dixon
            ,R.A.
  TITLE     Expressed Sequence Tags from the Samuel Roberts Noble Foundation
            Medicago truncatula stem library
  JOURNAL   Unpublished
COMMENT     On Jul 14, 2000 this sequence version replaced gi:9199881.
            Contact: Dixon RA
            Plant Biology Division
            The Samuel Roberts Noble Foundation
            2510 Sam Noble Parkway, Ardmore, OK 73402, USA
            Tel: 580 221 7302
            Fax: 580 221 7380
            Email: radixon@noble.org
            Medicago Genome Initiative accession: MGI:S:26688
            Insert Length: 674   Std Error: 0.00
            Plate: 086  row: F  column: 04
            Seq primer: TCACACAGGAAACAGCTATGAC.
FEATURES            Location/Qualifiers
     source         1. .515
                    /organism="Medicago truncatula"
                    /mol_type="mRNA"
                    /db_xref="taxon:3880"
```

/clone="NF086F04ST"
/tissue_type="stem"
/dev_stage="Pooled developmental"
/clone_lib="Developing stem"
/note="Vector: Lambda Zap; Contains a mixture of internodal stem segments"
BASE COUNT 143 a 89 c 127 g 156 t
ORIGIN

Alignment Scores:

Pred. No.:	7.41	Length:	515
Score:	9.00	Matches:	9
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	16.07%	Indels:	0
DB:	10	Gaps:	0

US-09-980-043-4 (1-56) x BE326104 (1-515)

```
Qy         21 AsnGlyGluLeuPheValAlaCysHis 29
              |||||||||||||||||||||||||||
Db        185 AATGGAGAGCTTTTTGTGGCTTGTCAT 211
```

RESULT 4
CA919485
LOCUS CA919485 558 bp mRNA linear EST 27-DEC-2002
DEFINITION EST637203 MTUS Medicago truncatula cDNA clone MTUS-14C9, mRNA sequence.
ACCESSION CA919485
VERSION CA919485.1 GI:27406415
KEYWORDS EST.
SOURCE Medicago truncatula (barrel medic)
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids ; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.
REFERENCE 1 (bases 1 to 558)
AUTHORS VandenBosch,K., Endre,G., Silverstein,K, Town,C.D., Van Aken,S., Utterback,T., Cheung,F. and Fraser,C.M.
TITLE The Medicago truncatula 6K unigene set: cDNA clones selected and re-arrayed from various libraries
JOURNAL Unpublished
COMMENT Contact: VandenBosch K
Department of Plant Biology
University of Minnesota
220 BioSci Center, 1445 Gortner Ave, St. Paul, MN 55108, USA
Tel: 612 624 2755
Fax: 612 625 1738
Email: kvandenb@cbs.umn.edu
TIGR sequence name: MTUAN14TV
More information is available at: www.medicago.org
Seq primer: (gtA AtA CgA CtC ACt AtA ggg C)
FEATURES Location/Qualifiers
source 1..558
/organism="Medicago truncatula"
/mol_type="mRNA"
/cultivar="A17"
/db_xref="taxon:3880"
/clone="MTUS-14C9"
/tissue_type="mixed tissues"
/dev_stage="various stages"
/lab_host="XLOLR"
/clone_lib="MTUS"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA was directionally ligated into the Unizap XR vector from Stratagene and packaged using Gigapack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-Zap phage using Ex-assist helper phage and propagated in XLOLR cells."
BASE COUNT 181 a 80 c 148 g 149 t
ORIGIN

Alignment Scores:

Pred. No.:	8.1	Length:	558
Score:	9.00	Matches:	9
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	16.07%	Indels:	0
DB:	14	Gaps:	0

US-09-980-043-4 (1-56) x CA919485 (1-558)

```
Qy         21 AsnGlyGluLeuPheValAlaCysHis 29
              |||||||||||||||||||||||||||
Db        123 AATGGAGAGCTTTTTGTGGCTTGTCAT 149
```

RESULT 5
BB334832
LOCUS BB334832 290 bp mRNA linear EST 11-JUL-2000
DEFINITION BB334832 RIKEN full-length enriched, 10 days neonate medulla oblongata Mus musculus cDNA clone B830028D15 3', mRNA sequence.
ACCESSION BB334832
VERSION BB334832.1 GI:9043595
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 290)
AUTHORS Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci ,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N., Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M. , Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y. , Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata ,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y. , Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya ,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I. , Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino ,M., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Konno,H., et al.)
JOURNAL Unpublished
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Thermostabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki ,Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for further details.
FEATURES Location/Qualifiers
source 1..290
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="B830028D15"
/tissue_type="medulla oblongata"

/dev_stage="10 days neonate"
/lab_host="DH10B"
/clone_lib="RIKEN full-length enriched, 10 days neonate medulla oblongata"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 459.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

SE COUNT 89 a 57 c 73 g 71 t
IGIN

ignment Scores:

ed. No.:	48.1	Length:	290
ore:	8.00	Matches:	8
rcent Similarity:	100.00%	Conservative:	0
st Local Similarity:	100.00%	Mismatches:	0
ery Match:	14.29%	Indels:	0
.	10	Gaps:	0

-09-980-043-4 (1-56) x BB334832 (1-290)

```
Qy         49 LysValCysLeuArgCysGlySer 56
              |||||||||||||||||||||||
Db         94 AAAGTATGCTTAAGATGTGGAAGT 117
```

SULT 6
701242
CUS BQ701242 318 bp mRNA linear EST 07-MAY-2003
FINITION NXSI_061_C04_F NXSI (Nsf Xylem Side wood Inclined) Pinus taeda cDNA clone NXSI_061_C04 5' similar to Arabidopsis thaliana sequence At5g17420 cellulose synthase catalytic subunit (IRX3) see http://mips.gsf.de/proj/thal/db/index.html, mRNA sequence.
CESSION BQ701242
RSION BQ701242.1 GI:21826558
YWORDS EST.
URCE Pinus taeda (loblolly pine)
ORGANISM Pinus taeda
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.
FERENCE 1 (bases 1 to 318)
AUTHORS Sederoff,R.
TITLE Molecular Basis of Wood Formation in the Pine Megagenome
JOURNAL Unpublished
MMENT Contact: Sederoff, Ron
Forest Biotechnology
North Carolina State University
840 Main Campus Drive, Centennial Campus, Campus Box 7247, Raleigh, NC 27695, USA
Tel: 919 515 7800
Fax: 919 515 7801
Email: ron_sederoff@ncsu.edu
Please see http://web.ahc.umn.edu/biodata/nsfpine/ for further information.
Seq primer: T3.
ATURES Location/Qualifiers
source 1. .318
/organism="Pinus taeda"
/mol_type="mRNA"
/strain="Coastal plain loblolly pine from North Carolina"
/db_xref="taxon:3352"
/clone="NXSI_061_C04"
/tissue_type="Xylem"
/cell_type="Side"
/dev_stage="Juvenile"
/lab_host="XL1-Blue"
/clone_lib="NXSI (Nsf Xylem Side wood Inclined)"
/note="Vector: BlueScript SK; Site_1: Eco RI; Site_2: XhoI ; The library is from early (spring) wood, taken from three six year old trees (three different genotypes), in the juvenile phase. These trees were induced to form side wood by bending to a 45 degree angle and tying them to the ground. Differentiating xylem was harvested from the sides of the inclined stems, and a mixture of all three genotypes was used for the library. oligo-dT primed cDNA was directionally cloned into the EcoRI-XhoI BlueScript SK vector (rms). NOTE: The sequences contain a 'cDNA adapter' between the EcoRI site and the start of the EST. The adapter sequence is 'AATTCGGCACGAG'."
BASE COUNT 82 a [illegible] 11 others
CRIGIN

Alignment Scores:

Pred. No.:	53.2	Length:	318
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	14.29%	Indels:	0
DB:	13	Gaps:	0

US-09-980-043-4 (1-56) x BQ702242 (1-318)

```
Qy         21 AsnGlyGluLeuPheValAlaCys 28
              ||||||||||.|||.||||||.|
Db        194 AATGGGGAACTATTCGTGGCCTGC 217
```

RESULT 7
CB702847/c
LOCUS CB702847 385 bp mRNA linear EST 10-APR-2003
DEFINITION AMGNNUC:NRWA3-00054-D8-A white adipose tiss (10469) Rattus norvegicus cDNA clone nrwa3-00054-d8 5', mRNA sequence.
ACCESSION CB702847
VERSION CB702847.1 GI:23759995
KEYWORDS EST.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 385)
AUTHORS Amgen EST Program.
TITLE Amgen Rat EST Program
JOURNAL Unpublished
COMMENT Contact: Dan Fitzpatrick
Amgen, Inc
One Amgen Center Drive, Thousand Oaks, CA 91320-1799, USA
Tel: 805 447-4861
Plate: 00054 row: d column: 8.
FEATURES Location/Qualifiers
source 1. .385
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/clone="nrwa3-00054-d8"
/tissue_type="adipose tiss"
/clone_lib="white adipose tiss (10469)"
/note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; white adipose tiss adult female Wistar rats, >250 grams"
BASE COUNT 56 a 125 c 81 g 67 t 56 others
ORIGIN

Alignment Scores:

Pred. No.:	65.7	Length:	385
Score:	8.00	Matches:	8

```
Percent Similarity:     100.00%        Conservative:  0
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            14.29%         Indels:        0
DB:                     14             Gaps:          0

US-09-980-043-4 (1-56) x CB702847 (1-385)

Qy         46 GluGlyArgLysValCysLeuArg 53
              |||||||||||||||||||||||
Db        177 GAGGGACGGAAAGTTTGCTTGCGG 154

RESULT 8
CB778315/c
LOCUS       CB778315                390 bp    mRNA    linear   EST 16-MAY-2003
DEFINITION  AMGNNUC:NRPI4-00058-E10-A W Rat pituitary (10472) Rattus norvegicus
            cDNA clone nrpi4-00058-e10 5', mRNA sequence.
ACCESSION   CB778315
VERSION     CB778315.1  GI:29866706
KEYWORDS    EST.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 390)
  AUTHORS   Amgen EST Program.
  TITLE     Amgen Rat EST Program
  JOURNAL   Unpublished
COMMENT     Contact: Dan Fitzpatrick
            Amgen, Inc
            One Amgen Center Drive, Thousand Oaks, CA 91320-1799, USA
            Tel: 805 447-4881
            Plate: 00058  row: e  column: 10
FEATURES             Location/Qualifiers
     source          1..390
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10116"
                     /clone="nrpi4-00058-e10"
                     /clone_lib="W Rat pituitary (10472)"
                     /note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; W Rat
                     pituitary adult female Wistar rat  vg insert size 2.1 kb"
BASE COUNT       62 a    147 c    102 g     79 t
ORIGIN

Alignment Scores:
Pred. No.:              66.6           Length:        390
Score:                  8.00           Matches:       8
Percent Similarity:     100.00%        Conservative:  0
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            14.29%         Indels:        0
DB:                     14             Gaps:          0

US-09-980-043-4 (1-56) x CB778315 (1-390)

Qy         46 GluGlyArgLysValCysLeuArg 53
              |||||||||||||||||||||||
Db        129 GAGGGACGGAAAGTTTGCTTGCGG 106

RESULT 9
AQ610092/c
LOCUS       AQ610092                396 bp    DNA     linear   GSS 15-JUN-1999
DEFINITION  HS_5092_A2_F09_SP6E RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=668 Col=18 Row=K, genomic survey sequence.
ACCESSION   AQ610092
VERSION     AQ610092.1  GI:5071368
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 396)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
   PUBMED   10449764
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 668  row: K  column: 18
            Seq primer: SP6
            Class: BAC ends
            High quality sequence stop: 396.
FEATURES             Location/Qualifiers
     source          1..396
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="Plate=668 Col=18 Row=K"
                     /sex="male"
                     /clone_lib="RPCI-11 Human Male BAC Library"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     Male blood DNA was isolated from one randomly chosen donor
                     and partially digested with a combination of EcoRI and
                     EcoRI Methylase. Size selected DNA was cloned into the
                     pBACe3.6 vector at EcoRI sites"
BASE COUNT      104 a     99 c     71 g    120 t      2 others
ORIGIN

Alignment Scores:
Pred. No.:              67.8           Length:        396
Score:                  8.00           Matches:       8
Percent Similarity:     100.00%        Conservative:  0
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            14.29%         Indels:        0
DB:                     28             Gaps:          0

US-09-980-043-4 (1-56) x AQ610092 (1-396)

Qy         46 GluGlyArgLysValCysLeuArg 53
              |||||||||||||||||||||||
Db         90 GAAGGCAGAAAAGTATGCTTGAGA 67

RESULT 10
CB768576/c
LOCUS       CB768576                403 bp    mRNA    linear   EST 16-MAY-2003
DEFINITION  AMGNNUC:CR3-00007-D12-A Colon rat 3 (10415) Rattus norvegicus cDNA
            clone cr3-00007-d12 5', mRNA sequence.
ACCESSION   CB768576
VERSION     CB768576.1  GI:29856967
KEYWORDS    EST.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 403)
  AUTHORS   Amgen EST Program.
  TITLE     Amgen Rat EST Program
  JOURNAL   Unpublished
COMMENT     Contact: Dan Fitzpatrick
            Amgen, Inc
```

```
          One Amgen Center Drive, Thousand Oaks, CA 91320-1799, USA
          Tel: 805 447-4881
          Plate: 00007  row: d  column: 12.
ATURES               Location/Qualifiers
    source           1. .403
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10116"
                     /clone="cr3-00007-d12"
                     /tissue_type="intestine, fetal"
                     /clone_lib="Colon rat 3 (10415)"
                     /note="Vector: pSPORT1; Site_1: Sal1; Site_2: Not1; Colon
                     rat 3 Fetal rat intestine"
SE COUNT        63 a    154 c    104 g     82 t
IGIN

ignment Scores:
ed. No.:              69.1          Length:        403
ore:                  8.00          Matches:       8
rcent Similarity:     100.00%       Conservative:  0
st Local Similarity:  100.00%       Mismatches:    0
ery Match:            14.29%        Indels:        0
:                     14            Gaps:          0

-09-980-043-4 (1-56) x CB768576 (1-403)

          46 GluGlyArgLysValCysLeuArg 53
             ||||||||||||||||||||||
         130 GAGGGACGGAAAGTTTGCTTGCGG 107

SULT 11
807973/c
CUS        CB807973                406 bp    mRNA    linear   EST 16-MAY-2003
FINITION   AMGNNUC:MRPE3-00085-E2-A placenta embryo D17 (10379) Rattus
           norvegicus cDNA clone mrpe3-00085-e2 5', mRNA sequence.
CESSION    CB807973
RSION      CB807973.1  GI:29927493
YWORDS     EST.
URCE       Rattus norvegicus (Norway rat)
ORGANISM   Rattus norvegicus
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
           Rattus.
FERENCE    1  (bases 1 to 406)
AUTHORS    Amgen EST Program.
TITLE      Amgen Rat EST Program
JOURNAL    Unpublished
OMMENT     Contact: Dan Fitzpatrick
           Amgen, Inc
           One Amgen Center Drive, Thousand Oaks, CA 91320-1799, USA
           Tel: 805 447-4881
           Plate: 00085  row: e  column: 2.
ATURES               Location/Qualifiers
    source           1. .406
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10116"
                     /clone="mrpe3-00085-e2"
                     /tissue_type="placenta embryo"
                     /clone_lib="placenta embryo D17 (10379)"
                     /note="Vector: pSPORT1; placenta embryo D17"
ASE COUNT       59 a    141 c     92 g     78 t     36 others
RIGIN

lignment Scores:
red. No.:             69.6          Length:        406
core:                 8.00          Matches:       8
ercent Similarity:    100.00%       Conservative:  0
est Local Similarity: 100.00%       Mismatches:    0
uery Match:           14.29%        Indels:        0
B:                    14            Gaps:          0

S-09-980-043-4 (1-56) x CB807973 (1-406)

Qy          46 GluGlyArgLysVa:CysLeuArg 53
               ||||||||||||||:|||| |||||
Db         148 GAGGGACGGAAAGT:TGCTTGCGG 125

RESULT 12
AQ443329/c
LOCUS       AQ443329                416 bp    DNA     linear   GSS 31-MAR-1999
DEFINITION  HS_2231_B2_D06_M? CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=2231 Col=12 Row=H, genomic survey
            sequence.
ACCESSION   AQ443329
VERSION     AQ443329.1  GI:4?54668
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metaz a; Chordata, Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates, Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 4 6)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,C., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
   PUBMED   10449764
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3487
            Email: jwallace@u.washington.edu
            Clones may be purchased from Research Genetics (info@resgen.com)
            BAC end Web Server: http://www.htsc.washington.edu
            Plate: 2231  row: H  column: 12
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 416.
FEATURES             Location/Qualifiers
     source          1. .416
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="Plate=2231 Col=12 Row=H"
                     /sex="male"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
BASE COUNT      119 a     94 c     68 g    135 t
ORIGIN

Alignment Scores:
Pred. No.:              71.5          Length:        416
Score:                  8.00          Matches:       8
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            14.29%        Indels:        0
DB:                     28            Gaps:          0

US-09-980-043-4 (1-56) x AQ443329 (1-416)

Qy          44 IleLysGluGlyArgLysValCys 51
               |||||||||||||||||||||||
Db         291 ATTAAGGAGGGAAGAAAAGTATGC 268

RESULT 13
BY574406
LOCUS       BY574406                437 bp    mRNA    linear   EST 15-DEC-2002
DEFINITION  BY574406 RIKEN full-length enriched, pooled tissues, spleen and
            thymus lymphocytes Mus musculus cDNA clone I6C0004E14 3', mRNA
            sequence.
```

```
ACCESSION   BY574406
VERSION     BY574406.1  GI:26909588
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 437)
  AUTHORS   Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
            Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I., Kiyosawa,H.
            , Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A., Schonbach,C.,
            Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Hume,D.A.,
            Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H., Batalov,S.,
            Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V., Chothia,C., Corbani
            ,L.E., Cousins,S., Dalla,E., Dragani,T.A., Fletcher,C.F., Forrest
            ,A., Frazer,K.S., Gaasterland,T., Gariboldi,M., Gissi,C., Godzik,A.
            , Gough,J., Grimmond,S., Gustincich,S., Hirokawa,N., Jackson,I.J.,
            Jarvis,E.D., Kanai,A., Kawaji,H., Kawasawa,Y., Kedzierski,R.M.,
            King,B.L., Konagaya,A., Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons
            ,P.A., Maglott,D.R., Maltais,L., Marchionni,L., McKenzie,L., Miki
            ,H., Nagashima,T., Numata,K., Okido,T., Pavan,W.J., Pertea,G.,
            Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D.,
            Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring
            ,B.Z., Ringwald,M., Sandelin,A., Schneider,C., Semple,C.A., Setou
            ,M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale
            ,R.D., Tomita,M., Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y.,
            Watanabe,Y., Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa
            ,M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A.,
            Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura
            ,M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.
            , Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii
            ,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata
            ,K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander
            ,E.S., Rogers,J., Birney,E. and Hayashizaki,Y.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
  MEDLINE   22354683
   PUBMED   12466851
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hirozane
            ,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H., Miyazaki,A.
            , Murata,M., Nakamura,M., Nomura,K., Numazaki,R., Ohno,M., Sakai,K.
            , Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami
            ,M., Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y. Direct
            Submission
             Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences Mamm. Genome. 12, 671-677 (2001)
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. 10 (10), 1617-1630 (2000)
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
             cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
             Tissues were provided by Hiromitsu Nakauchi ( Dept. of Immunology
            Institute of Basic Medical Sciences University of Tsukuba 1-1-1
            Tennodai,Tsukuba,Ibaraki 305-8578 ) whose assistance we gratefully
            acknowledge.
             Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
FEATURES            Location/Qualifiers
     source         1. .437
                    /organism="Mus musculus"
                    /mol_type="mRNA"
                    /db_xref="taxon:10090"
                    /clone="I6C0004E14"
                    /clone_lib="RIKEN full-length enriched, pooled tissues,
                    spleen and thymus lymphocytes"
                    /note="pooled tissues: (tissue_type=thymus
                    ,cell_type=lymphocyte), (tissue_type=spleen
                    ,cell_type=lymphocyte)"
BASE COUNT      111 a    102 c    118 g    104 t      2 others
ORIGIN

Alignment Scores:
Pred. No.:                75.5          Length:          437
Score:                    8.00          Matches:         8
Percent Similarity:       100.00%       Conservative:    0
Best Local Similarity:    100.00%       Mismatches:      0
Query Match:              14.29%        Indels:          0
DB:                       13            Gaps:            0

US-09-980-043-4 (1-56) x BY574406 (1-437)

Qy         22 GlyGluLeuPheValAlaCysHis 29
              ||||||||||||||||||||||||
Db        273 GGGGAGCTGTTTGTTGCTTGCCAT 296

RESULT 14
CB001824/c
LOCUS       CB001824                 470 bp    mRNA    linear   EST 10-JAN-2003
DEFINITION  VVB009E02_125110 An expressed sequence tag database for abiotic
            stressed leaves of Vitis vinifera var. Chardonnay Vitis vinifera
            cDNA clone VVB009E02 5, mRNA sequence.
ACCESSION   CB001824
VERSION     CB001824.1  GI:27579129
KEYWORDS    EST.
SOURCE      Vitis vinifera
  ORGANISM  Vitis vinifera
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; Vitaceae; Vitis.
REFERENCE   1  (bases 1 to 470)
  AUTHORS   Cramer,G.R. and Cushman,J.C.
  TITLE     An expressed sequence tag database for abiotic stressed leaves of
            Vitis vinifera var. Chardonnay
  JOURNAL   Unpublished
COMMENT     Contact: Cushman JC
            Department of Biochemistry
            University of Nevada
            MS200, Reno, NV 89557-0014, USA
            Tel: 775-784-1918
            Fax: 775-784-1650
            Email: jcushman@unr.edu
            PCR PRimers
            FORWARD: T3 20mer
            BACKWARD: T7 21mer
            Plate: 0  row: E  column: 02
            Seq primer: T3 20mer
            High quality sequence stop: 470.
FEATURES            Location/Qualifiers
     source         1. .470
                    /organism="Vitis vinifera"
                    /mol_type="mRNA"
                    /db_xref="taxon:29760"
                    /clone="VVB009E02"
                    /tissue_type="leaf"
                    /dev_stage="juvenile and adult"
                    /clone_lib="An expressed sequence tag database for abiotic
                    stressed leaves of Vitis vinifera var. Chardonnay"
```

/note="Vector: Lambda Uni-Zap XR, Bluescript SK-; Site_1: EcoRI; Site_2: XhoI; Library construction was performed according to Stratagene's recommended protocol for the Lambda UniZapXR vector and cDNA synthesis kit."

SE COUNT 119 a 118 c 97 g 136 t
IGIN

ignment Scores:

ed. No.:	81.8	Length:	470
ore:	8.00	Matches:	8
rcent Similarity:	100.00%	Conservative:	0
st Local Similarity:	100.00%	Mismatches:	0
ery Match:	14.29%	Indels:	0
:	14	Gaps:	0

-09-980-043-4 (1-56) x CB001824 (1-470)

```
Qy         42 PheGluIleLysGluGlyArgLys 49
              |||||||||||||||||||||||
Db        407 TTTGAAATTAAAGAGGGGAGGAAG 384
```

SULT 15
661271/c

```
CUS       AQ661271                500 bp    DNA     linear   GSS 23-JUN-1999
FINITION  Sheared DNA-5C19.TR Sheared DNA Trypanosoma brucei genomic clone
          Sheared DNA-5C19, genomic survey sequence.
CESSION   AQ661271
RSION     AQ661271.1  GI:5169039
YWORDS    GSS.
URCE      Trypanosoma brucei
ORGANISM  Trypanosoma brucei
          Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
          Trypanosoma.
FERENCE   1  (bases 1 to 500)
AUTHORS   El-Sayed,N., Zhao,S., Zhao,H., Gill,S., Suh,E., Malek,J., Fujii,C.,
          Gerrard,C., Leech,V., de Jong,P., Ullu,E., Melville,S., Donelson,J.
          , Fraser,C. and Adams,M.
TITLE     Determination of clone end sequences from Trypanosoma brucei GUTat
          10.1 sheared DNA library
JOURNAL   Unpublished
MMENT     Other_GSSs: Sheared DNA-5C19.TF
          Contact: Najib M. El-Sayed
          Department of Eukaryotic Genomics
          The Institute for Genomic Research
          9712 Medical Center Dr., Rockville, MD 20850, USA
          Tel: 301 838 0200
          Fax: 301 838 0208
          Email: nelsayed@tigr.org
          Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared
          DNA library constructed at TIGR. Clones will be available for
          distribution through ATCC. Sheared DNA end sequences search page:
          http://www.tigr.org/tdb/mdb/tbdb/.
          Seq primer: M13-Reverse
          Class: shotgun.
ATURES              Location/Qualifiers
   source           1. .500
                    /organism="Trypanosoma brucei"
                    /mol_type="genomic DNA"
                    /strain="TREU927/4 GUTat 10.1"
                    /db_xref="taxon:5691"
                    /clone="Sheared DNA-5C19"
                    /clone_lib="Sheared DNA"
                    /note="Vector: pUC18; Site_1: SmaI; Constructed at The
                    Institute for Genomic Research (TIGR), Rockville, MD.
                    Genomic DNA isolated from a cloned population of
                    Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically
                    sheared to give a tight size distribution (approx 2 kb).
                    The v + i method used for the library construction is
                    described in detail in Smith, H.O. and Venter, J.C.
                    (Making small insert libraries for whole genome shotgun
                    sequencing projects.  In Genome Sequencing: A Practical
                    Approach, eds. M. Vaudin and B. Barell, Oxford University
                    Press, 1999)."
```

BASE COUNT 140 a [illegible] 98 g [illegible]
ORIGIN

Alignment Scores:

Pred. No.:	87.6	Length:	500
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	14.29%	Indels:	0
DB:	28	Gaps:	0

US-09-980-043-4 (1-56) x AQ[illegible]271 (1-500)

```
Qy         43 GluIleLysGluGl ArgLysVal 56
              ||||||||||||| ||||||
Db        466 GAAATAAAGGAGG AGAAAGTA 443
```

Search completed: Novem[illegible], 84.18.58
Job time : 340.263 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 29, 2003, 18:30:59 ; Search time 3.84526 Seconds
(without alignments)
3758.118 Million cell updates/sec

Title: US-09-980-043-4
Perfect score: 56
Sequence: 1 MMESGAPICHTCGEQVGHDA.....KSCFEPEIKEGRKVCLRCGS 56

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 830525 seqs, 258052604 residues

Word size : 0

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SPTREMBL_23:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	56	100.0	56	10	Q9LLJ2	Q9llj2 populus tre
2	44	78.6	978	10	O81368	O81368 populus tre
3	8	14.3	326	10	Q39895	Q39895 glycine max
4	8	14.3	375	5	Q8I211	Q8i211 plasmodium
5	8	14.3	958	10	Q9SN37	Q9sn37 arabidopsis
6	8	14.3	985	10	Q8LPK5	Q8lpk5 arabidopsis
7	8	14.3	985	10	Q9C5Z8	Q9c5z8 arabidopsis
8	7	12.5	138	4	Q8NA59	Q8na59 homo sapien
9	7	12.5	154	5	P91131	P91131 caenorhabdi
10	7	12.5	192	10	Q94JP7	Q94jp7 thespesia l
11	7	12.5	192	10	Q94JP6	Q94jp6 thespesia t
12	7	12.5	192	10	Q94JP8	Q94jp8 hampea appe
13	7	12.5	193	10	Q94JP3	Q94jp3 cienfuegosi
14	7	12.5	193	10	Q94JP5	Q94jp5 thespesia p
15	7	12.5	193	10	Q9M693	Q9m693 gossypioide
16	7	12.5	193	10	Q94JP9	Q94jp9 kokia dryna
17	7	12.5	197	10	Q93WX1	Q93wx1 gossypium l
18	7	12.5	197	10	Q93WX2	Q93wx2 gossypium s
19	7	12.5	197	10	Q93WX3	Q93wx3 gossypium a
20	7	12.5	197	10	Q93WX0	Q93wx0 gossypium b
21	7	12.5	197	10	Q8GV67	Q8gv67 gossypium g
22	7	12.5	197	10	Q8GV66	Q8gv66 gossypium d
23	7	12.5	197	10	Q8GV65	Q8gv65 gossypium s
24	7	12.5	197	10	Q8GV64	Q8gv64 gossypium t
25	7	12.5	197	10	Q8GV63	Q8gv63 gossypium t
26	7	12.5	205	10	Q9SYX4	Q9syx4 gossypium h
27	7	12.5	205	10	Q9XFC6	Q9xfc6 gossypium r
28	7	12.5	205	10	Q9XFC5	Q9xfc5 gossypium r
29	7	12.5	205	10	Q9XFC4	Q9xfc4 gossypium h
30	7	12.5	205	10	Q9SYX5	Q9syx5 gossypium h
31	7	12.5	310	2	Q8GCA8	Q8gca8 bacillus me
32	7	12.5	312	5	O96116	O96116 plasmodium
33	7	12.5	321	10	Q8GRY7	Q8gry7 lotus japon
34	7	12.5	322	10	O04234	O04234 vicia faba
35	7	12.5	322	10	Q39896	Q39896 glycine max
36	7	12.5	342	5	Q9VY77	Q9vy77 drosophila
37	7	12.5	360	5	Q8ID02	Q8id02 plasmodium
38	7	12.5	385	17	Q8PXZ4	Q8pxz4 methanosarc
39	7	12.5	452	16	Q8RET8	Q8ret8 fusobacteri
40	7	12.5	974	10	P93155	P93155 gossypium h
41	7	12.5	974	10	Q8W1W0	Q8w1w0 gossypium h
42	7	12.5	1129	10	Q8LK26	Q8lk26 mesotaenium
43	7	12.5	2895	10	O65551	O65551 arabidopsis
44	6	10.7	51	17	Q8U0V6	Q8u0v6 pyrococcus
45	6	10.7	79	2	Q8KMV1	Q8kmv1 escherichia

ALIGNMENTS

RESULT 1
Q9LLJ2
ID Q9LLJ2 PRELIMINARY; PRT; 56 AA.
AC Q9LLJ2;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Cellulose synthase (Fragment).
GN CESA.
OS Populus tremuloides (Quaking aspen).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Malpighiales; Salicaceae; Populus.
OX NCBI_TaxID=3693;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Xylem;
RX MEDLINE=20344632; PubMed=10886769;
RA Wu L., Joshi C.P., Chiang V.L.;
RT "A xylem-specific cellulose synthase gene from aspen (Populus
RT tremuloides) is responsive to mechanical stress.";
RL Plant J. 22:495-502(2000).
DR EMBL; AF197911; AAF91155.1; -.
FT NON_TER 56 56
SQ SEQUENCE 56 AA; 6160 MW; 11BFFF8C37328402 CRC64;

Query Match 100.0%; Score 56; DB 10; Length 56;
Best Local Similarity 100.0%; Pred. No. 6.6e-55;
Matches 56; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MMESGAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEPEIKEGRKVCLRCGS 56
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 MMESGAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEPEIKEGRKVCLRCGS 56

RESULT 2
O81368
ID O81368 PRELIMINARY; PRT; 978 AA.
AC O81368;

01-NOV-1998 (TrEMBLrel. 08, Created)
01-JAN-1999 (TrEMBLrel. 09, Last sequence update)
01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
Secondary xylem cellulose synthase.
CELA.
Populus tremuloides (Quaking aspen).
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
eurosids I; Malpighiales; Salicaceae; Populus.
NCBI_TaxID=3693;
[1]
SEQUENCE FROM N.A.
TISSUE=Secondary xylem;
MEDLINE=20344632; PubMed=10886769;
Wu L., Joshi C.P., Chiang V.L.;
"A xylem-specific cellulose synthase gene from aspen (Populus
tremuloides) is responsive to mechanical stress.";
Plant J. 22:495-502(2000).
EMBL; AF072131; AAD03417.1; -.
InterPro; IPR005150; Cellulose_synt.
Pfam; PF03552; Cellulose_synt; 1.
SEQUENCE 978 AA; 110279 MW; C351B0E4975C6B8A CRC64;

Query Match 78.6%; Score 44; DB 10; Length 978;
Best Local Similarity 100.0%; Pred. No. 2e-40;
Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MMESGAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEI 44
||||||||||||||||||||||||||||||||||||||||||||
Db 1 MMESGAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEI 44

RESULT 3
Q9895
ID Q39895 PRELIMINARY; PRT; 326 AA.
AC Q39895;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE TGACG-motif binding factor.
GN Glycine max (Soybean).
OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
OX NCBI_TaxID=3847;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Williams; TISSUE=Hypocotyl;
RX MEDLINE=98388650; PubMed=9721678;
RA Cheong Y.H., Yoo C.M., Park J.M., Ryu G.R., Goekjian V.H., Nagao R.T.,
RA Key J.L., Cho M.J., Hong J.C.;
RT "STF1 is a novel TGACG-binding factor with a zinc-finger motif and a
RT bZIP domain which heterodimerizes with GBF proteins.";
RL Plant J. 15:199-209(1998).
CC -!- SIMILARITY: BELONGS TO THE BZIP FAMILY.
DR EMBL; L28003; AAC05017.1; -.
DR TRANSFAC; T02969; -.
DR InterPro; IPR004827; TF_bZIP.
DR InterPro; IPR001841; Znf_ring.
DR Pfam; PF00170; bZIP; 1.
DR SMART; SM00338; BRLZ; 1.
DR SMART; SM00184; RING; 1.
DR PROSITE; PS00036; BZIP_BASIC; 1.
DR PROSITE; PS50089; ZF_RING_2; 1.
KW DNA-binding; Nuclear protein.
SQ SEQUENCE 326 AA; 35488 MW; 0BF960434DC19AC5 CRC64;

Query Match 14.3%; Score 8; DB 10; Length 326;
Best Local Similarity 100.0%; Pred. No. 1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 24 LFVACHEC 31
||||||||
Db 54 LFVACHEC 61

RESULT 4
Q8I211
ID Q8I211 PRELIMINARY PRT; 375 AA.
AC Q8I211;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE RIFIN.
GN RIF.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22255708; PubMed=12368865;
RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,
RA Mungall K., Bowman S., Tivey R., Baker S., Barron A., Brooks K.,
RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,
RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,
RA Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,
RA Feltwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,
RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,
RA Humphray S., Jagels K., James K.D., Johnson D., Kerhornou A.,
RA Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Lennard N.,
RA Line A., Maddison M., Mclean J., Mooney P., Moule S., Murphy L.,
RA Oliver K., Ormond D., Price C., Quail M.A., Rabbinowitsch E.,
RA Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,
RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,
RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,
RA Sulston J.E., Craig A., Newbold C., Barrell B.G;
RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13.";
RL Nature 419:527-531(2002).
DR EMBL; AL034557; CAD49109.2; -.
SQ SEQUENCE 375 AA; 41779 MW; BE24F7509F3EDCA9 CRC64;

Query Match 14.3%; Score 8; DB 5; Length 375;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 49 KVCLRCGS 56
||||||||
Db 142 KVCLRCGS 149

RESULT 5
Q9SN37
ID Q9SN37 PRELIMINARY; PRT; 958 AA.
AC Q9SN37;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Cellulose synthase-like protein.
GN F28A21.190 OR AT4G18780
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Mueller M.W., Muendlein A., Felber R., Bancroft I.,
RA Mewes H.W., Mayer K.F.X., Lemcke K., Schueller C.;
RL Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Mueller M.W., Muendlein A., Felber R., Mewes H.W., Lemcke K.,

```
RA   Mayer K.F.X.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   EU Arabidopsis sequencing project;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL035526; CAB37463.1; -.
DR   EMBL; AL161549; CAB78880.1; -.
DR   InterPro; IPR005150; Cellulose_synt.
DR   Pfam; PF03552; Cellulose_synt; 1.
SQ   SEQUENCE   958 AA;  108520 MW;  B3836EF4BC697502 CRC64;

  Query Match             14.3%;  Score 8;  DB 10;  Length 958;
  Best Local Similarity   100.0%;  Pred. No. 2.7;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         25 FVACHECS 32
              ||||||||
Db         25 FVACHECS 32


RESULT 6
Q8LPK5
ID   Q8LPK5      PRELIMINARY;      PRT;   985 AA.
AC   Q8LPK5;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Cellulose synthase-like protein.
GN   AT4G18780.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Nguyen M., Karlin-Neumann G., Southwick A., Lam B., Miranda M.,
RA   Palm C.J., Bowser L., Jones T., Bank J., Carninci P., Chen H.,
RA   Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA   Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA   Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
RA   Ecker J., Theologis A., Davis R.W.;
RL   Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY099636; AAM20487.1; -.
DR   InterPro; IPR005150; Cellulose_synt
DR   Pfam; PF03552; Cellulose_synt; 1.
SQ   SEQUENCE   985 AA;  111521 MW;  D299F357625768E0 CRC64;

  Query Match             14.3%;  Score 8;  DB 10;  Length 985;
  Best Local Similarity   100.0%;  Pred. No. 2.8;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         25 FVACHECS 32
              ||||||||
Db         25 FVACHECS 32


RESULT 7
Q9C5Z8
ID   Q9C5Z8      PRELIMINARY;      PRT;   985 AA.
AC   Q9C5Z8;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Cellulose synthase catalytic subunit.
GN   IRX1.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21066939; PubMed=11148295;
RA   Taylor N.G., Laurie S., Turner S.R.;
RT   "Multiple Cellulose Synthase Catalytic Subunits Are Required for
RT   Cellulose Synthesis in Arabidopsis.";
RL   Plant Cell 12:2529-2540(2000).
DR   EMBL; AF267742; AAK08700.1; -.
DR   InterPro; IPR005150; Cellulose_synt.
DR   Pfam; PF03552; Cellulose_synt; 1.
SQ   SEQUENCE   985 AA;  111470 MW;  693C701B7B69AB25 CRC64;

  Query Match             14.3%;  Score 8;  DB 10;  Length 985;
  Best Local Similarity   100.0%;  Pred. No. 2.8;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         25 FVACHECS 32
              ||||||||
Db         25 FVACHECS 32


RESULT 8
Q8NA59
ID   Q8NA59      PRELIMINARY;      PRT;   138 AA.
AC   Q8NA59;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hypothetical protein FLJ35816.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Testis;
RA   Oshima A., Takahashi-Fujii A., Tanase T., Imose N., Takeuchi K.,
RA   Arita M., Musashino K., Yuuki H., Hara H., Sugiyama T., Irie R.,
RA   Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA   Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA   Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M.,
RA   Murakawa K., Kanehori K., Sugiyama A., Kawakami B., Suzuki Y.,
RA   Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK093135; BAC04069.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   138 AA;  14436 MW;  D42023E97E3A2F23 CRC64;

  Query Match             12.5%;  Score 7;  DB 4;  Length 138;
  Best Local Similarity   100.0%;  Pred. No. 6.4;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 ESGAPIC 9
              |||||||
Db         39 ESGAPIC 45


RESULT 9
P91131
ID   P91131      PRELIMINARY;      PRT;   154 AA.
AC   P91131;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical 17.8 kDa protein.
GN   C32E8.1.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
```

```
   MEDLINE=99069613; PubMed=9851916;
   None;
   "Genome sequence of the nematode C. elegans: a platform for
   investigating biology. The C. elegans Sequencing Consortium.";
   Science 282:2012-2018(1998).
   [2]
   SEQUENCE FROM N.A.
   STRAIN=Bristol N2;
   Gattung S.;
   "The sequence of C. elegans cosmid C32E8.";
   Submitted (FEB-1997) to the EMBL/GenBank/DDBJ databases.
   [3]
   SEQUENCE FROM N.A.
   STRAIN=Bristol N2;
   Waterston R.;
   "Direct Submission.";
   Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
   -!- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
   EMBL; U88308; AAB42321.1; -.
   WormPep; C32E8.1; CE08525.
   InterPro; IPR001841; Znf_ring.
   Pfam; PF00097; zf-C3HC4; 1.
   SMART; SM00184; RING; 1.
   PROSITE; PS00518; ZF_RING_1; 1.
   PROSITE; PS50089; ZF_RING_2; 1.
   Hypothetical protein; Metal-binding; Zinc; Zinc-finger.
   SEQUENCE   154 AA;  17788 MW;  0B600EBCAD4F4881 CRC64;

Query Match             12.5%;  Score 7;  DB 5;  Length 154;
Best Local Similarity   100.0%;  Pred. No. 7;
Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

         14 EQVGHDA 20
            |||||||
         13 EQVGHDA 19


SULT 10
4JP7
   Q94JP7      PRELIMINARY;      PRT;   192 AA.
   Q94JP7;
   01-DEC-2001 (TrEMBLrel. 19, Created)
   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
   Cellulose synthase A1 (Fragment).
   CESA1.
   Thespesia lampas.
   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
   eurosids II; Malvales; Malvaceae; Malvoideae; Thespesia.
   NCBI_TaxID=47619;
   [1]
   SEQUENCE FROM N.A.
   Wendel J.F., Cronn R.C., Johnston S., Price J.;
   "Feast and Famine in Plant Genomes.";
   Genetica 0:0-0(2001).
   EMBL; AF376043; AAK57497.1; -.
   InterPro; IPR005150; Cellulose_synt.
   Pfam; PF03552; Cellulose_synt; 1.
   NON_TER       1       1
   NON_TER     192     192
   SEQUENCE   192 AA;  21718 MW;  6E3BC63B7B9F93CC CRC64;

Query Match             12.5%;  Score 7;  DB 10;  Length 192;
Best Local Similarity   100.0%;  Pred. No. 8.5;
Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

         25 FVACHEC 31
            |||||||
         11 FVACHEC 17


SULT 11
```

```
Q94JP6
ID   Q94JP6      PRELIMINARY;      PRT;   192 AA.
AC   Q94JP6;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Cellulose synthase A1 (Fragment).
GN   CESA1.
OS   Thespesia thespesioides
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Malvales; Malvaceae; Malvoideae; Thespesia.
OX   NCBI_TaxID=47611;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Wendel J.F., Cronn R.C., Johnston S., Price J.;
RT   "Feast and Famine in Plant Genomes.";
RL   Genetica 0:0-0(2001).
DR   EMBL; AF376044; AAK57498.1; -.
DR   InterPro; IPR005150; Cellulose_synt
DR   Pfam; PF03552; Cellulose_synt; 1.
FT   NON_TER       1       1
FT   NON_TER     192     192
SQ   SEQUENCE   192 AA;  21718 MW;  6E3BC63B7B9F93CC CRC64;

  Query Match             12.5%;  Score 7;  DB 10;  Length 192;
  Best Local Similarity   100.0%;  Pred. No. 8.5;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         25 FVACHEC 31
              |||||||
Db         11 FVACHEC 17


RESULT 12
Q94JP8
ID   Q94JP8      PRELIMINARY;      PRT;   192 AA.
AC   Q94JP8;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Cellulose synthase A1 (Fragment).
GN   CESA1.
OS   Hampea appendiculata.
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Malvales; Malvaceae; Malvoideae; Hampea.
OX   NCBI_TaxID=47613;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Wendel J.F., Cronn R.C., Johnston S., Price J.;
RT   "Feast and Famine in Plant Genomes.";
RL   Genetica 0:0-0(2001).
DR   EMBL; AF376042; AAK57496.1; -.
DR   InterPro; IPR005150; Cellulose_synt
DR   Pfam; PF03552; Cellulose_synt; 1.
FT   NON_TER       1       1
FT   NON_TER     192     192
SQ   SEQUENCE   192 AA;  21679 MW;  9CCE398D61A7D431 CRC64;

  Query Match             12.5%;  Score 7;  DB 10;  Length 192;
  Best Local Similarity   100.0%;  Pred. No. 8.5;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         25 FVACHEC 31
              |||||||
Db         11 FVACHEC 17


RESULT 13
Q94JP3
ID   Q94JP3      PRELIMINARY;      PRT;   193 AA.
AC   Q94JP3;
```

```
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Cellulose synthase A1 (Fragment).
GN   CESA1.
OS   Cienfuegosia tripartita.
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Malvales; Malvaceae; Malvoideae; Cienfuegosia.
OX   NCBI_TaxID=47608;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Wendel J.F., Cronn R.C., Johnston S., Price J.;
RT   "Feast and Famine in Plant Genomes.";
RL   Genetica 0:0-0(2001).
DR   EMBL; AF376047; AAK57501.1; -.
DR   InterPro; IPR005150; Cellulose_synt.
DR   Pfam; PF03552; Cellulose_synt; 1.
FT   NON_TER       1      1
FT   NON_TER     193    193
SQ   SEQUENCE   193 AA;  21625 MW;  D2A3F149FA4A8FA2 CRC64;

  Query Match             12.5%;  Score 7;  DB 10;  Length 193;
  Best Local Similarity   100.0%;  Pred. No. 8.5;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         25 FVACHEC 31
              |||||||
Db         11 FVACHEC 17


RESULT 14
Q94JP5
ID   Q94JP5      PRELIMINARY;      PRT;   193 AA.
AC   Q94JP5;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update
DE   Cellulose synthase A1 (Fragment)
GN   CESA1.
OS   Thespesia populnea (Alamo blanco).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Malvales; Malvaceae; Malvoideae; Thespesia.
OX   NCBI_TaxID=3638;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Wendel J.F., Cronn R.C., Johnston S., Price J.;
RT   "Feast and Famine in Plant Genomes.";
RL   Genetica 0:0-0(2001).
DR   EMBL; AF376045; AAK57499.1; -.
DR   InterPro; IPR005150; Cellulose_synt
DR   Pfam; PF03552; Cellulose_synt; 1
FT   NON_TER       1      1
FT   NON_TER     193    193
SQ   SEQUENCE   193 AA;  21699 MW;  BD1C0855E0284DD5 CRC64;

  Query Match             12.5%;  Score 7;  DB 10;  Length 193;
  Best Local Similarity   100.0%;  Pred. No. 8.5;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         25 FVACHEC 31
              |||||||
Db         11 FVACHEC 17


RESULT 15
Q9M693
ID   Q9M693      PRELIMINARY;      PRT;   191 AA.
AC   Q9M693;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Cellulose synthase A1 (Fragment).
OS   Gossypioides kirkii.
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Malvales; Malvaceae; Malvoideae; Gossypioides.
OX   NCBI_TaxID=47615;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20056257; PubMed=10588718;
RA   Cronn R.C., Small R.L., Wendel J.F.;
RT   "Duplicated genes evolve independently after polyploid formation in
RT   cotton.";
RL   Proc. Natl. Acad. Sci. U.S.A. 96:14406-14411(1999).
DR   EMBL; AF201886; AAF31409.1; -.
DR   InterPro; IPR005150; Cellulose_synt.
DR   Pfam; PF03552; Cellulose_synt; 1.
FT   NON_TER       1      1
FT   NON_TER     193    193
SQ   SEQUENCE   193 AA;  21727 MW;  2609E8109A71DAA0 CRC64;

  Query Match             12.5%;  Score 7;  DB 10;  Length 193;
  Best Local Similarity   100.0%;  Pred. No. 8.5;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         25 FVACHEC 31
              |||||||
Db         12 FVACHEC 18


Search completed: October 29, 2003, 18:35:48
Job time : 5.84526 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 29, 2003, 18:31:34 ; Search time 1.19149 Seconds
(without alignments)
2210.255 Million cell updates/sec

Title: US-09-980-043-4
Perfect score: 56
Sequence: 1 MMESGAPICHTCGEQVGHDA.....KSCFEFEIKEGRKVCLRCGS 56

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 127863 seqs, 47026705 residues

Word size : 0

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database: SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID		Description
1	7	12.5	44	1	Y135_HELPY	O24948	helicobacte
2	6	10.7	159	1	DUT_CANAL	P43058	candida alb
3	6	10.7	229	1	GLC8_YEAST	P41818	saccharomyc
4	6	10.7	229	1	GPMA_LISMC	Q8y571	listeria mo
5	6	10.7	247	1	RS3_TREPA	O83225	treponema p
6	6	10.7	261	1	AAC4_SALSP	P08988	salmonella
7	6	10.7	263	1	IF2A_SULTO	Q973g0	sulfolobus
8	6	10.7	270	1	CWFG_SCHPO	Q9p7c5	schizosacch
9	6	10.7	324	1	HEM3_PAEMA	O69110	paenibacill
10	6	10.7	359	1	LLCD_SYNY3	Q08871	synechocyst
11	6	10.7	394	1	TGT_DEIRA	Q9rrb5	deinococcus
12	6	10.7	395	1	METC_ECOLI	P06721	escherichia
13	6	10.7	395	1	METC_SALTY	P18949	salmonella
14	6	10.7	424	1	EXON_NPVOP	P24081	orgyia pseu
15	6	10.7	444	1	RBL_OSMCI	P33536	osmunda cin
16	6	10.7	448	1	XYLA_BACLI	P77832	bacillus li
17	6	10.7	450	1	NRL1_HUMAN	Q9bu40	homo sapien
18	6	10.7	533	1	VMCB_LAMBD	P03710	bacteriopha
19	6	10.7	551	1	RN27_HUMAN	Q9bzr9	homo sapien
20	6	10.7	551	1	RN27_MOUSE	Q99pj2	mus musculu
21	6	10.7	557	1	YFK7_YEAST	P43609	saccharomyc
22	6	10.7	578	1	DFA2_SYNEL	Q8dj55	synechococc
23	6	10.7	657	1	MCPA_CAUCR	Q00986	caulobacter
24	6	10.7	682	1	CYG2_RAT	P22717	rattus norv
25	6	10.7	688	1	YP22_YEAST	Q08968	saccharomyc
26	6	10.7	719	1	NRP1_YEAST	P32770	saccharomyc
27	6	10.7	721	1	RIR1_MYCPN	P78027	mycoplasma
28	6	10.7	765	1	S23A_HUMAN	Q15436	homo sapien
29	6	10.7	765	1	S23A_MOUSE	Q01405	mus musculu
30	6	10.7	864	1	41_HUMAN	P11171	homo sapien
31	6	10.7	898	1	PODK_RHIME	Q59754	rhizobium m
32	6	10.7	1020	1	CONT_MOUSE	P12960	mus musculu
33	6	10.7	1021	1	CONT_RAT	Q63198	rattus norv
34	6	10.7	1056	1	CARB_PYRFU	Q8u085	pyrococcus
35	6	10.7	1059	1	CAPU_DROME	Q24120	drosophila
36	6	10.7	1180	1	PC12_MOUSE	O55134	mus musculu
37	6	10.7	1407	1	RPOC_ECOLI	P00577	escherichia
38	6	10.7	2332	1	POLG_FMDVA	P03308	f genome po
39	6	10.7	2332	1	POLG_FMDVO	P03305	f genome po
40	6	10.7	2333	1	POLG_FMDV1	P03306	f genome po
41	6	10.7	2336	1	POLG_FMDVZ	P49303	f genome po
42	5	8.9	45	1	Y5K6_SSV1	P20198	sulfolobus
43	5	8.9	51	1	RL40_PYRHO	O74010	pyrococcus
44	5	8.9	54	1	RS14_PYRAE	Q8zvw1	pyrobaculum
45	5	8.9	55	1	RL15_LACLC	P27145	lactococcus

ALIGNMENTS

RESULT 1
Y135_HELPY
ID Y135_HELPY STANDARD; PRT; 44 AA.
AC O24948;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Hypothetical protein HP0135/JHP0123 precursor.
GN HP0135 OR JHP0123.
OS Helicobacter pylori (Campylobacter pylori), and
OS Helicobacter pylori J99 (Campylobacter pylori J99).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=210, 85963;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=26695 / ATCC 700392;
RX MEDLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA McKenney K., FitzGerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
RT pylori.";
RL Nature 388:539-547(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=J99;
RX MEDLINE=99120557; PubMed=9923682;
RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA Trust T.J.;
RT "Genomic sequence comparison of two unrelated isolates of the human
RT gastric pathogen Helicobacter pylori.";
RL Nature 397:176-180(1999).
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; AE000535; AAD07212.1; -.
DR EMBL; AE001451; AAD05700.1; -.
DR PIR; G64536; G64536.
DR TIGR; HP0135; -.
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; UNKNOWN_1.

```
   Hypothetical protein; Signal; Complete proteome.
   SIGNAL        1     16       POTENTIAL.
   CHAIN        17     44       HYPOTHETICAL PROTEIN HP0135/JHP0123.
)  SEQUENCE   44 AA;  5023 MW;  868B773CB3807A7A CRC64;

Query Match             12.5%;  Score 7;  DB 1;  Length 44;
Best Local Similarity   100.0%;  Pred. No. 0.68;
Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

         24 LFVACHE 30
            |||||||
         13 LFVACHE 19



SULT 2
T_CANAL
   DUT_CANAL      STANDARD;      PRT;   159 AA.
   P43058;
   01-NOV-1995 (Rel. 32, Created)
   01-NOV-1995 (Rel. 32, Last sequence update)
   28-FEB-2003 (Rel. 41, Last annotation update)
   Deoxyuridine 5'-triphosphate nucleotidohydrolase (EC 3.6.1.23)
   (dUTPase) (dUTP pyrophosphatase).
   DUT1.
   Candida albicans (Yeast).
   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
   Saccharomycetales; mitosporic Saccharomycetales; Candida.
   NCBI_TaxID=5476;
   [1]
   SEQUENCE FROM N.A.
   STRAIN=WO-1;
   MEDLINE=95179795; PubMed=7874734;
   McIntosh E., Looser J., Haynes R.H., Pearlman R.E.;
   "MluI site-dependent transcriptional regulation of the Candida
   albicans dUTPase gene.";
   Curr. Genet. 26:415-421(1994).
   -!- FUNCTION: THIS ENZYME IS INVOLVED IN NUCLEOTIDE METABOLISM: IT
       PRODUCES DUMP, THE IMMEDIATE PRECURSOR OF THYMIDINE NUCLEOTIDES
       AND IT DECREASES THE INTRACELLULAR CONCENTRATION OF DUTP SO THAT
       URACIL CANNOT BE INCORPORATED INTO DNA.
   -!- CATALYTIC ACTIVITY: dUTP + H(2)O = dUMP + diphosphate.
   -!- PATHWAY: De novo synthesis of thymidylate.
   -!- SIMILARITY: BELONGS TO THE DUTPASE FAMILY.
   ----------------------------------------------------------------------
   This SWISS-PROT entry is copyright. It is produced through a collaboration
   between the Swiss Institute of Bioinformatics and the EMBL outstation -
   the European Bioinformatics Institute. There are no restrictions on its
   use by non-profit institutions as long as its content is in no way
   modified and this statement is not removed. Usage by and for commercial
   entities requires a license agreement (See http://www.isb-sib.ch/announce/
   or send an email to license@isb-sib.ch).
   ----------------------------------------------------------------------
   EMBL; X77925; CAA54897.1; -.
   PIR; S51498; S51498.
   HSSP; P06968; 1EUW.
   InterPro; IPR001428; DeoxyUTPase.
   Pfam; PF00692; dUTPase; 1.
   ProDom; PD000946; dUTPase_2; 1.
   TIGRFAMs; TIGR00576; dut; 1.
   Hydrolase; Nucleotide metabolism.
   SEQUENCE   159 AA;  16945 MW;  982C016F1D31013A CRC64;

Query Match             10.7%;  Score 6;  DB 1;  Length 159;
Best Local Similarity   100.0%;  Pred. No. 23;
Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

         42 FEIKEG 47
            ||||||
        114 FEIKEG 119



SULT 3
C8_YEAST
```

```
ID   GLC8_YEAST     STANDARD;      PRT;   229 AA.
AC   P41818;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   GLC8 protein.
GN   GLC8 OR YMR311C OR YM9954.03C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94200592; PubMed=8150278;
RA   Cannon J.F., Pringle J.R., Fiechter A., Khalil M.;
RT   "Characterization of glycogen-deficient glc mutants of Saccharomyces
RT   cerevisiae.";
RL   Genetics 136:485-503(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / AB972;
RX   PubMed=9169872;
RA   Bowman S., Churcher C.M., Badcock K., Brown D., Chillingworth T.,
RA   Connor R., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S.,
RA   Jagels K., Lye G., Moule S., Odell C., Pearson D., Rajandream M.A.,
RA   Rice P., Skelton J., Walsh S., Whitehead S., Barrell B.G.;
RT   "The nucleotide sequence of Saccharomyces cerevisiae chromosome
RT   XIII.";
RL   Nature 387:90-93(1997).
CC   -!- FUNCTION: MODULATOR OF GLC7 TYPE-1 PROTEIN PHOSPHATASE.
CC   -!- SIMILARITY: SOME, TO MAMMALIAN PROTEIN PHOSPHATASE INHIBITOR 2.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; L22000; AAA53673.1; -.
DR   EMBL; Z54141; CAA90829.1; -.
DR   PIR; S47859; S47859.
DR   SGD; S0004928; GLC8.
DR   GO; GO:0005737; C:cytoplasm; IDA.
DR   GO; GO:0008047; F:enzyme activator activity; IDA.
KW   Phosphorylation; Glycogen metabolism.
FT   MOD_RES     118    118       PHOSPHORYLATION (BY GSK3) (BY
FT                                SIMILARITY).
SQ   SEQUENCE   229 AA;  26649 MW;  D4017CA8B3B089C2 CRC64;

  Query Match             10.7%;  Score 6;  DB 1;  Length 229;
  Best Local Similarity   100.0%;  Pred. No. 31;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          41 EFEIKE 46
               ||||||
Db         163 EFEIKE 168


RESULT 4
GPMA_LISMO
ID   GPMA_LISMO     STANDARD;      PRT;   229 AA.
AC   Q8Y571;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   2,3-bisphosphoglycerate-dependent phosphoglycerate mutase (EC 5.4.2.1)
DE   (Phosphoglyceromutase) (PGAM) (BPG-dependent PGAM) (dPGM).
GN   GPMA OR LMO2205.
OS   Listeria monocytogenes.
OC   Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX   NCBI_TaxID=1639;
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=EGD-e / Serovar 1/2a;
RX   MEDLINE=21537279; PubMed=11679669;
RA   Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA   Baquero F., Berche P., Bloecker H., Brandt P., Chakraborty T.,
RA   Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA   Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA   Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA   Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA   Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkat G.,
RA   Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA   Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA   Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA   Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT   "Comparative genomics of Listeria species.";
RL   Science 294:849-852(2001).
CC   -!- FUNCTION: Catalyzes the interconversion of 2-phosphoglycerate and
CC       3-phosphoglycerate (By similarity).
CC   -!- CATALYTIC ACTIVITY: 2-phospho-D-glycerate + 2,3-diphosphoglycerate
CC       = 3-phospho-D-glycerate + 2,3-diphosphoglycerate.
CC   -!- PATHWAY: Glycolysis.
CC   -!- SIMILARITY: BELONGS TO THE PHOSPHOGLYCERATE MUTASE FAMILY. BPG-
CC       DEPENDENT PGAM SUBFAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AL591982; CAD00283.1; -.
DR   PIR; AE1350; AE1350.
DR   ListiList; LMO02205; -.
DR   HAMAP; MF_01039; -; 1.
DR   InterPro; IPR001345; PG/BPGM_mutase.
DR   InterPro; IPR005952; Phosphogly_mut1.
DR   Pfam; PF00300; PGAM; 1.
DR   TIGRFAMs; TIGR01258; pgm_1; 1.
DR   PROSITE; PS00175; PG_MUTASE; 1.
KW   Isomerase; Glycolysis; Complete proteome.
FT   ACT_SITE      8      8       PHOSPHOHISTIDINE INTERMEDIATE
FT                                (BY SIMILARITY).
FT   ACT_SITE     59     59       REQUIRED FOR BINDING CARBOXYL GROUP OF
FT                                PHOSPHOGLYCERATES (BY SIMILARITY).
FT   ACT_SITE    181    181       BY SIMILARITY.
SQ   SEQUENCE   229 AA;  26417 MW;  3073F6E8CEBCEFE9 CRC64;

  Query Match             10.7%;  Score 6;  DB 1;  Length 229;
  Best Local Similarity   100.0%;  Pred. No. 31;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         43 EIKEGR 48
              ||||||
Db        169 EIKEGR 174


RESULT 5
RS3_TREPA
ID   RS3_TREPA      STANDARD;      PRT;   247 AA.
AC   O83225;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   30S ribosomal protein S3.
GN   RPSC OR TP0195.
OS   Treponema pallidum.
OC   Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.
OX   NCBI_TaxID=160;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Nichols;
RX   MEDLINE=98332770; PubMed=9665876;
RA   Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
RA   Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA   Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA   Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
RA   McDonald L., Artiach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA   Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA   Venter J.C.;
RT   "Complete genome sequence of Treponema pallidum, the syphilis
RT   spirochete.";
RL   Science 281:375-388(1998).
CC   -!- FUNCTION: Binds the lower part of the 30S subunit head. Binds mRNA
CC       in the 70S ribosome, positioning it for translation (By
CC       similarity).
CC   -!- SUBUNIT: Part of the 30S ribosomal subunit. Forms a tight complex
CC       with proteins S10 and S14 (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE S3P FAMILY OF RIBOSOMAL PROTEINS.
CC   -!- SIMILARITY: Contains 1 KH type-2 domain.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AE001202; AAC65180.1; -.
DR   PIR; E71355; E71355.
DR   TIGR; TP0195; -.
DR   HAMAP; MF_01309; -; 1.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004044; KH_TYPE_2.
DR   InterPro; IPR001351; Ribosomal_S3.
DR   InterPro; IPR005704; S3_bact.
DR   Pfam; PF00013; KH; 1.
DR   Pfam; PF00189; Ribosomal_S3_C; 1.
DR   Pfam; PF00417; Ribosomal_S3_N; 1.
DR   SMART; SM00322; KH; 1.
DR   TIGRFAMs; TIGR01009; rpsC_bact; 1.
DR   PROSITE; PS50823; KH_TYPE_2; 1.
DR   PROSITE; PS00548; RIBOSOMAL_S3; 1.
KW   Ribosomal protein; RNA-binding; rRNA-binding; Complete proteome.
FT   DOMAIN       39    109       KH TYPE-2.
SQ   SEQUENCE   247 AA;  27772 MW;  232F8A74C9DE3EC4 CRC64;

  Query Match             10.7%;  Score 6;  DB 1;  Length 247;
  Best Local Similarity   100.0%;  Pred. No. 33;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         43 EIKEGR 48
              ||||||
Db        169 EIKEGR 174


RESULT 6
AAC4_SALSP
ID   AAC4_SALSP     STANDARD;      PRT;   261 AA.
AC   P08988;
DT   01-NOV-1988 (Rel. 09, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Aminoglycoside N(3')-acetyltransferase IV (EC 2.3.1.81) (ACC(3)-IV)
DE   (Aminocyclitol 3-N-acetyltransferase type IV).
GN   AACC4.
OS   Salmonella sp.
OG   Plasmid pWP7b.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Salmonella.
OX   NCBI_TaxID=599;
RN   [1]
RP   SEQUENCE FROM N.A.
```

MEDLINE=84092826; PubMed=6318050;
Braeu B., Pilz U., Piepersberg W.;
"Genes for gentamicin-(3)-N-acetyltransferases III and IV: I. Nucleotide sequence of the AAC(3)-IV gene and possible involvement of an IS140 element in its expression.";
Mol. Gen. Genet. 193:179-187(1984).
[2]
REVISIONS.
Piepersberg W.;
Submitted (SEP-1985) to the EMBL/GenBank/DDBJ databases.
-!- FUNCTION: RESISTANCE TO ANTIBIOTICS CONTAINING THE 2-DEOXY-STREPTAMINE RING INCLUDING GENTAMICIN, KANAMYCIN, TOBRAMYCIN, NEOMYCIN AND APRAMYCIN.
-!- CATALYTIC ACTIVITY: Acetyl-CoA + a 2-deoxystreptamine antibiotic = CoA + N(3')-acetyl-2-deoxystreptamine antibiotic.
-!- SIMILARITY: TO OTHER AAC(3) PROTEINS.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).

EMBL; X01385; CAA25642.1; -.
PIR; S09647; S09647.
InterPro; IPR003679; Antibiotic_NAT.
Pfam; PF02522; Antibiotic_NAT; 1.
Antibiotic resistance; Transferase; Acyltransferase; Plasmid.
SEQUENCE 261 AA; 28491 MW; 7E8CBBFC6DFC111E CRC64;

Query Match 10.7%; Score 6; DB 1; Length 261;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

16 VGHDAN 21
||||||
145 VGHDAN 150

SULT 7
2A_SULTO
IF2A_SULTO STANDARD; PRT; 263 AA.
Q973G0;
28-FEB-2003 (Rel. 41, Created)
28-FEB-2003 (Rel. 41, Last sequence update)
28-FEB-2003 (Rel. 41, Last annotation update)
Probable translation initiation factor 2 alpha subunit (eIF-2-alpha).
EIF2A OR ST0941.
Sulfolobus tokodaii.
Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae; Sulfolobus.
NCBI_TaxID=111955;
[1]
SEQUENCE FROM N.A.
STRAIN=JCM 10545 / 7;
MEDLINE=21456156; PubMed=11572479;
Kawarabayasi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H., Hosoyama A., Fukui S., Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y., Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A., Oshima T., Kikuchi H.;
"Complete genome sequence of an aerobic thermoacidophilic Crenarchaeon, Sulfolobus tokodaii strain7.";
DNA Res. 8:123-140(2001).
-!- FUNCTION: eIF-2 functions in the early steps of protein synthesis by forming a ternary complex with GTP and initiator tRNA (By similarity).
-!- SUBUNIT: Heterotrimer composed of an alpha, a beta, and a gamma chain (By similarity).
-!- SIMILARITY: Belongs to the eIF-2-alpha family.
-!- SIMILARITY: Contains 1 S1 motif domain.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).

DR EMBL; AP000984; BAB65955.1; -.
DR HAMAP; MF_00231; -; 1.
DR InterPro; IPR003029; S1
DR Pfam; PF00575; S1; 1.
DR SMART; SM00316; S1; 1.
DR PROSITE; PS50126; S1; 1
KW Initiation factor; Protein biosynthesis; RNA binding;
KW Complete proteome.
FT DOMAIN 12 83 S1 MOTIF
SQ SEQUENCE 263 AA; 30[illegible] MW; AC[illegible]6497 CRC64;

Query Match 10.7%; Score 6; DB 1; Length 263;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 44 IKEGRK 49
||||||
Db 58 IKEGRK 63

RESULT 8
CWFG_SCHPO
ID CWFG_SCHPO STANDARD; PRT; 270 AA.
AC Q9P7C5; Q9UT15;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cell cycle control protein cwf16.
GN CWF16 OR SPAC9.13C OR SPAPJ735.01C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes; Schizosaccharomycetales; Schizosaccharomycetaceae; Schizosaccharomyces.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A., Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S., Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M., Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A., Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G., Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K., James K., Jones L., Jones M., Leather S., McDonald S., McLean J., Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C., Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E., Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S., Skelton J., Simmonds M., Squares R., Squares S., Stevens K., Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S., Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B., Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H., Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M., Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B., Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S., Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M., Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G., Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J., Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L., Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J., Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe.";

```
RL   Nature 415:871-880(2002).
RN   [2]
RP   IDENTIFICATION.
RX   MEDLINE=21881936; PubMed=11884590;
RA   Ohi M.D., Link A.J., Ren L., Jennings J.L., McDonald W.H., Gould K.L.;
RT   "Proteomics analysis reveals stable multiprotein complexes in both
RT   fission and budding yeasts containing Myb-related Cdc5p/Cef1p, novel
RT   pre-mRNA splicing factors, and snRNAs.";
RL   Mol. Cell. Biol. 22:2011-2024(2002).
CC   -!- FUNCTION: Involved in mRNA splicing where it associates with cdc5
CC       and the other cwf proteins as part of the spliceosome.
CC   -!- SUBUNIT: Interacts with cdc5.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AL163031; CAB86262.1; -.
DR   EMBL; AL121764; CAB57431.1; -.
DR   GeneDB_SPombe; SPAC9.13c; -.
DR   Pfam; PF04502; DUF572; 1.
KW   Nuclear protein; mRNA processing; mRNA splicing.
SQ   SEQUENCE   270 AA;  31653 MW;  3A403699FCC655CE CRC64;

  Query Match             10.7%;  Score 6;  DB 1;  Length 270;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          9 CHTCGE 14
              ||||||
Db         48 CHTCGE 53


RESULT 9
HEM3_PAEMA
ID   HEM3_PAEMA     STANDARD;      PRT;   324 AA.
AC   O69110;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Porphobilinogen deaminase (EC 4.3.1.8) (PBG) (Hydroxymethylbilane
DE   synthase) (HMBS) (Pre-uroporphyrinogen synthase).
GN   HEMC.
OS   Paenibacillus macerans (Bacillus macerans).
OC   Bacteria; Firmicutes; Bacillales; Paenibacillaceae; Paenibacillus.
OX   NCBI_TaxID=44252;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DSM 24;
RA   Johansson P., Hederstedt L.;
RT   "Organization of genes for tetrapyrrole biosynthesis in Gram-positive
RT   bacteria.";
RL   Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: TETRAPOLYMERIZATION OF THE MONOPYRROLE PBG INTO THE
CC       HYDROXYMETHYLBILANE PREUROPORPHYRINOGEN IN SEVERAL DISCRETE STEPS.
CC   -!- CATALYTIC ACTIVITY: 4 porphobilinogen + H(2)O =
CC       hydroxymethylbilane + 4 NH(3).
CC   -!- COFACTOR: COVALENTLY BINDS A DIPYRROMETHANE COFACTOR TO WHICH THE
CC       PORPHOBILINOGEN SUBUNITS ARE ADDED (BY SIMILARITY).
CC   -!- PATHWAY: Porphyrin biosynthesis by the C5 pathway; fourth step.
CC   -!- SUBUNIT: Monomer (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE HMBS FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF064061; AAC18587.1; -.
DR   HSSP; P06983; 1PDA.
DR   HAMAP; MF_00260; atypical; 1.
DR   InterPro; IPR000860; Porphobil_deam.
DR   Pfam; PF01379; Porphobil_deam; 1.
DR   Pfam; PF03900; Porphobil_deamC; 1.
DR   PRINTS; PR00151; PORPHBDMNASE.
DR   ProDom; PD002745; Porphobil_deam; 1.
DR   TIGRFAMs; TIGR00212; hemC; 2.
DR   PROSITE; PS00533; PORPHOBILINOGEN_DEAM; 1.
KW   Porphyrin biosynthesis; Lyase.
FT   DOMAIN      261    279       INSERT.
FT   BINDING     246    246       PYRROMETHANE COFACTOR (BY SIMILARITY).
SQ   SEQUENCE   324 AA;  34903 MW;  9F674AA6865B2F85 CRC64;

  Query Match             10.7%;  Score 6;  DB 1;  Length 324;
  Best Local Similarity   100.0%;  Pred. No. 41;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         40 FEFEIK 45
              ||||||
Db         38 FEFEIK 43


RESULT 10
LLCD_SYNY3
ID   LLCD_SYNY3     STANDARD;      PRT;   359 AA.
AC   Q08871;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Linoleoyl-CoA desaturase (EC 1.14.19.3) (Delta(6)-desaturase).
GN   DES6 OR SLL0262.
OS   Synechocystis sp. (strain PCC 6803).
OC   Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX   NCBI_TaxID=1148;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93283633; PubMed=8389613;
RA   Reddy A.S., Nuccio M.L., Gross L.M., Thomas T.L.;
RT   "Isolation of a delta 6-desaturase gene from the cyanobacterium
RT   Synechocystis sp. strain PCC 6803 by gain-of-function expression in
RT   Anabaena sp. strain PCC 7120.";
RL   Plant Mol. Biol. 22:293-300(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97061201; PubMed=8905231;
RA   Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA   Miyajima N., Hirosawa M., Sugiura M., Sasamoto S., Kimura T.,
RA   Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K.,
RA   Okumura S., Shimpo S., Takeuchi C., Wada T., Watanabe A.,
RA   Yamada M., Yasuda M., Tabata S.;
RT   "Sequence analysis of the genome of the unicellular cyanobacterium
RT   Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT   entire genome and assignment of potential protein-coding regions.";
RL   DNA Res. 3:109-136(1996).
CC   -!- CATALYTIC ACTIVITY: Linoleoyl-CoA + AH(2) + O(2) = gamma-
CC       linolenoyl-CoA + A + 2 H(2)O.
CC   -!- COFACTOR: IRON.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L11421; AAA27286.1; -.
DR   EMBL; D90914; BAA18502.1; -.
```

InterPro; IPR005804; FA_desat_fam.
Pfam; PF00487; FA_desaturase; 1.
ProDom; PD001081; FA_desat_fam; 1.
Oxidoreductase; Iron; Complete proteome.
SEQUENCE 359 AA; 41425 MW; 33FB165AEB98C05F CRC64;

Query Match 10.7%; Score 6; DB 1; Length 359;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

16 VGHDAN 21
||||||
86 VGHDAN 91

SULT 11
T_DEIRA
TGT_DEIRA STANDARD; PRT; 394 AA.
Q9RRB5;
28-FEB-2003 (Rel. 41, Created)
28-FEB-2003 (Rel. 41, Last sequence update)
28-FEB-2003 (Rel. 41, Last annotation update)
Queuine tRNA-ribosyltransferase (EC 2.4.2.29) (tRNA-guanine transglycosylase) (Guanine insertion enzyme).
TGT OR DR2578.
Deinococcus radiodurans.
Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales; Deinococcaceae; Deinococcus.
NCBI_TaxID=1299;
[1]
SEQUENCE FROM N.A.
STRAIN=R1 / ATCC 13939 / DSM 20539 / NCIB 9279;
MEDLINE=20036896; PubMed=10567266;
White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D., Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L., Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M., Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C., Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D., Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C., Fraser C.M.;
"Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.";
Science 286:1571-1577(1999).
-!- FUNCTION: Exchanges the guanine residue with 7-aminomethyl-7-deazaguanine in tRNAS with GU(N) anticodons (tRNA-Asp, -Asn, -His and -Tyr). After this exchange, a cyclopentendiol moiety is attached to the 7-aminomethyl group of 7-deazaguanine, resulting in the hypermodified nucleoside queuosine (Q) (7-(((4,5-cis-dihydroxy-2-cyclopenten-1-yl)amino)methyl)-7-deazaguanosine) (By similarity).
-!- CATALYTIC ACTIVITY: tRNA guanine + queuine = tRNA queuine + guanine.
-!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
-!- SIMILARITY: BELONGS TO THE QUEUINE TRNA-RIBOSYLTRANSFERASE FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).

EMBL; AE002087; AAF12118.1; -.
PIR; D75256; D75256.
HSSP; P28720; 1PUD.
TIGR; DR2578; -.
HAMAP; MF_00168; -; 1.
InterPro; IPR004803; QtRNA_ribo_trans.
InterPro; IPR002616; tRNA_ribo_trans.
Pfam; PF01702; TGT; 1.
TIGRFAMs; TIGR00430; Q_tRNA_tgt; 1.
TIGRFAMs; TIGR00449; tgt_general; 1.

KW Queuosine biosynthesis; Transferase; Glycosyltransferase;
KW tRNA processing; Zinc; Complete proteome.
FT ACT_SITE 100 100 BY SIMILARITY.
FT ACT_SITE 273 273 BY SIMILARITY.
FT METAL 310 310 ZINC (BY SIMILARITY).
FT METAL 312 312 ZINC (BY SIMILARITY).
FT METAL 315 315 ZINC (BY SIMILARITY).
FT METAL 341 341 ZINC (BY SIMILARITY).
SQ SEQUENCE 394 AA; 438 MW; 74EAB943EDC3247F CRC64;

Query Match 10.7%; Score 6; DB 1; Length 394;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 40 FEFEIK 45
||||||
Db 11 FEFEIK 16

RESULT 12
METC_ECOLI
ID METC_ECOLI STANDARD; PRT; 395 AA.
AC P06721;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cystathionine beta-lyase (EC 4.4.1.8) (CBL) (Beta-cystathionase) (Cysteine lyase).
GN METC OR B3008.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86149241; PubMed=3513164;
RA Belfaiza J., Parsot C., Martel A., de la Tour C.B., Margarita D., Cohen G.N., Saint-Girons I.;
RT "Evolution in biosynthetic pathways: two enzymes catalyzing consecutive steps in methionine biosynthesis originate from a common ancestor and possess a similar regulatory region.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:867-871(1986).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V., Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F., Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J., Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (1.83 ANGSTROMS)
RX MEDLINE=96428687; PubMed=8831789;
RA Clausen T., Huber R., Laber B., Pohlenz H.-D., Messerschmidt A.;
RT "Crystal structure of the pyridoxal-5'-phosphate dependent cystathionine beta-lyase from Escherichia coli at 1.83 A.";
RL J. Mol. Biol. 262:202-224(1996).
CC -!- CATALYTIC ACTIVITY: Cystathionine + H(2)O = L-homocysteine + NH(3) + pyruvate.
CC -!- COFACTOR: Pyridoxal phosphate.
CC -!- PATHWAY: Methionine biosynthesis; third step.
CC -!- SUBUNIT: Homotetramer.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to the trans sulfuration enzymes family.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/

```
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M12858; AAA24158.1; -.
DR   EMBL; U28377; AAA69175.1; ALT_INIT.
DR   EMBL; AE000383; AAC76044.1; -.
DR   PIR; A25153; WZECCB.
DR   PDB; 1CL1; 11-NOV-98.
DR   PDB; 1CL2; 14-OCT-98.
DR   ECO2DBASE; G043.0; 6TH EDITION.
DR   EcoGene; EG10583; metC.
DR   InterPro; IPR006233; Cys_b_lyase_bac.
DR   InterPro; IPR000277; Cys_Met_Meta_PP.
DR   Pfam; PF01053; Cys_Met_Meta_PP; 1.
DR   TIGRFAMs; TIGR01324; cysta_beta_ly_B; 1.
DR   PROSITE; PS00868; CYS_MET_METAB_PP; 1.
KW   Methionine biosynthesis; Lyase; Pyridoxal phosphate; 3D-structure;
KW   Complete proteome.
FT   BINDING     210    210       PYRIDOXAL PHOSPHATE.
FT   HELIX         6     12
FT   TURN         13     14
FT   HELIX        17     20
FT   TURN         21     22
FT   STRAND       23     23
FT   STRAND       28     28
FT   STRAND       34     34
FT   HELIX        39     47
FT   TURN         48     52
FT   TURN         57     59
FT   HELIX        62     75
FT   TURN         76     76
FT   STRAND       79     83
FT   HELIX        86     95
FT   TURN         96     97
FT   TURN        100    101
FT   STRAND      103    107
FT   TURN        108    109
FT   HELIX       112    120
FT   TURN        121    122
FT   HELIX       123    125
FT   TURN        126    126
FT   STRAND      128    132
FT   TURN        134    135
FT   HELIX       137    143
FT   TURN        146    147
FT   STRAND      148    154
FT   TURN        158    160
FT   HELIX       166    176
FT   TURN        178    179
FT   STRAND      181    185
FT   TURN        187    192
FT   HELIX       196    199
FT   TURN        200    203
FT   STRAND      203    207
FT   TURN        208    213
FT   STRAND      221    225
FT   TURN        227    229
FT   HELIX       230    239
FT   TURN        240    241
FT   HELIX       246    256
FT   TURN        257    258
FT   HELIX       259    278
FT   TURN        279    279
FT   TURN        281    282
FT   STRAND      283    287
FT   TURN        289    290
FT   TURN        292    293
FT   TURN        295    296
FT   HELIX       297    303
FT   STRAND      310    315
FT   HELIX       321    328
FT   TURN        329    330
FT   STRAND      335    336
FT   STRAND      346    350
FT   HELIX       352    357
FT   TURN        358    358
FT   TURN        360    361
FT   STRAND      370    374
FT   HELIX       380    393
FT   TURN        394    395
SQ   SEQUENCE   395 AA;  43212 MW;  FD3308423D6427C6 CRC64;

  Query Match             10.7%;  Score 6;  DB 1;  Length 395;
  Best Local Similarity   100.0%;  Pred. No. 48;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          20 ANGELF 25
               ||||||
Db          50 ANGELF 55


RESULT 13
METC_SALTY
ID   METC_SALTY     STANDARD;      PRT;   395 AA.
AC   P18949;
DT   01-NOV-1990 (Rel. 16, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Cystathionine beta-lyase (EC 4.4.1.8) (CBL) (Beta-cystathionase)
DE   (Cysteine lyase).
GN   METC OR STM3161.
OS   Salmonella typhimurium.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Salmonella.
OX   NCBI_TaxID=602;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LT2;
RX   MEDLINE=89281484; PubMed=2659970;
RA   Park Y.M., Stauffer G.V.;
RT   "DNA sequence of the metC gene and its flanking regions from
RT   Salmonella typhimurium LT2 and homology with the corresponding
RT   sequence of Escherichia coli.";
RL   Mol. Gen. Genet. 216:164-169(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX   MEDLINE=21534948; PubMed=11677609;
RA   McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA   Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA   Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA   Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA   Waterston R., Wilson R.K.;
RT   "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT   LT2.";
RL   Nature 413:852-856(2001).
CC   -!- CATALYTIC ACTIVITY: Cystathionine + H(2)O = L-homocysteine + NH(3)
CC       + pyruvate.
CC   -!- COFACTOR: Pyridoxal phosphate.
CC   -!- PATHWAY: Methionine biosynthesis; third step.
CC   -!- SUBUNIT: Homotetramer.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: Belongs to the trans-sulfuration enzymes family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE008845; AAL22035.1; -.
DR   PIR; JV0020; JV0020.
DR   HSSP; P06721; 1CL2.
DR   StyGene; SG10223; metC.
DR   InterPro; IPR006233; Cys_b_lyase_bac.
```

```
   InterPro; IPR000277; Cys_Met_Meta_PP.
   Pfam; PF01053; Cys_Met_Meta_PP; 1.
   TIGRFAMs; TIGR01324; cysta_beta_ly_B; 1.
   PROSITE; PS00868; CYS_MET_METAB_PP; 1.
   Methionine biosynthesis; Lyase; Pyridoxal phosphate;
   Complete proteome.
   BINDING     210     210       PYRIDOXAL PHOSPHATE.
   CONFLICT    195     195       K -> E (IN REF. 1).
   CONFLICT    200     200       D -> H (IN REF. 1).
   SEQUENCE   395 AA;  42872 MW;  38B0A41A1A063447 CRC64;

Query Match             10.7%;  Score 6;  DB 1;  Length 395;
Best Local Similarity   100.0%;  Pred. No. 48;
Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

        20 ANGELF 25
           ||||||
        50 ANGELF 55



SULT 14
ON_NPVOP
   EXON_NPVOP     STANDARD;      PRT;   424 AA.
   P24081;
   01-MAR-1992 (Rel. 21, Created)
   01-MAR-1992 (Rel. 21, Last sequence update)
   01-NOV-1997 (Rel. 35, Last annotation update)
   Putative alkaline exonuclease (EC 3.1.11.-).
   ALK-EXO.
   Orgyia pseudotsugata multicapsid polyhedrosis virus (OpMNPV).
   Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
   Nucleopolyhedrovirus.
   NCBI_TaxID=164623;
   [1]
   SEQUENCE FROM N.A.
   MEDLINE=89293088; PubMed=2661722;
   Gombart A.F., Blissard G.W., Rohrmann G.F.;
   "Characterization of the genetic organization of the HindIII M region
   of the multicapsid nuclear polyhedrosis virus of Orgyia pseudotsugata
   reveals major differences among baculoviruses.";
   J. Gen. Virol. 70:1815-1828(1989).
   [2]
   SEQUENCE FROM N.A.
   MEDLINE=97271300; PubMed=9126251;
   Ahrens C.H., Russell R.R., Funk C.J., Evans J., Harwood S.,
   Rohrmann G.F.;
   "The sequence of the Orgyia pseudotsugata multinucleocapsid nuclear
   polyhedrosis virus genome.";
   Virology 229:381-399(1997).
   -!- SIMILARITY: TO HERPES VIRUSES ALKALINE EXONUCLEASE.
   -!- SIMILARITY: 53% IDENTITY WITH THE ACMNPV ALK-EXO.
   ---------------------------------------------------------------------
   This SWISS-PROT entry is copyright. It is produced through a collaboration
   between the Swiss Institute of Bioinformatics and the EMBL outstation -
   the European Bioinformatics Institute. There are no restrictions on its
   use by non-profit institutions as long as its content is in no way
   modified and this statement is not removed. Usage by and for commercial
   entities requires a license agreement (See http://www.isb-sib.ch/announce/
   or send an email to license@isb-sib.ch).
   ---------------------------------------------------------------------
   EMBL; D13796; BAA02953.1; -.
   EMBL; D13929; BAA03031.1; -.
   EMBL; U75930; AAC59130.1; -.
   PIR; E30857; E30857.
   InterPro; IPR001616; Herpes_alk_exo.
   Pfam; PF01771; Herpes_alk_exo; 1.
   Hydrolase; Nuclease; Exonuclease; Late protein.
   SEQUENCE   424 AA;  47350 MW;  57C8EBBCF1C046AF CRC64;

Query Match             10.7%;  Score 6;  DB 1;  Length 424;
Best Local Similarity   100.0%;  Pred. No. 51;
Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          20 ANGELF 25
               ||||||
Db         104 ANGELF 109


RESULT 15
RBL_OSMCI
ID   RBL_OSMCI      STANDARD;      PRT;   444 AA.
AC   P33536;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Ribulose bisphosphate carboxylase large chain (EC 4.1.1.39) (RuBisCO
DE   large subunit) (Fragment).
GN   RBCL.
OS   Osmunda cinnamomea (Cinnamon fern).
OG   Chloroplast.
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Filicophyta; Filicopsida; Filicales; Osmundaceae; Osmunda.
OX   NCBI_TaxID=3284;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94111142; PubMed=8283479;
RA   Hasebe M., Ito M., Kofuji R., Ueda K., Iwatsuki K.;
RT   "Phylogenetic relationships of ferns deduced from rbcL gene
RT   sequence.";
RL   J. Mol. Evol. 37:476-482(1993).
CC   -!- FUNCTION: RuBisCO catalyzes two reactions: the carboxylation of D-
CC       ribulose 1,5-bisphosphate, the primary event in photosynthetic
CC       carbon dioxide fixation, as well as the oxidative fragmentation of
CC       the pentose substrate in the photorespiration process. Both
CC       reactions occur simultaneously and in competition at the same
CC       active site.
CC   -!- CATALYTIC ACTIVITY: D-ribulose 1,5-bisphosphate + CO(2) = 2 3-
CC       phospho-D-glycerate.
CC   -!- CATALYTIC ACTIVITY: D-ribulose 1,5-bisphosphate + O(2) =
CC       3-phospho-D-glycerate + 2-phosphoglycolate.
CC   -!- SUBUNIT: 8 large chains + 8 small chains.
CC   -!- SUBCELLULAR LOCATION: Chloroplast.
CC   -!- SIMILARITY: BELONGS TO THE RUBISCO LARGE CHAIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D14882; BAA03601.1; -.
DR   InterPro; IPR000685; RuBisCO_large.
DR   Pfam; PF00016; RuBisCO_large; 1.
DR   Pfam; PF02788; RuBisCO_large_N; 1.
DR   PROSITE; PS00157; RUBISCO_LARGE; 1.
KW   Photosynthesis; Carbon dioxide fixation; Photorespiration;
KW   Lyase; Oxidoreductase; Monooxygenase; Chloroplast.
FT   NON_TER       1       1
FT   ACT_SITE    192     192       BINDING OF CO(2) ACTIVATES THE ENZYME
FT   NON_TER     444     444
SQ   SEQUENCE   444 AA;  49025 MW;  2DA313FACC7F41FC CRC64;

  Query Match             10.7%;  Score 6;  DB 1;  Length 444;
  Best Local Similarity   100.0%;  Pred. No. 53;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy           3 ESGAPI 8
                ||||||
Db         250 ESGAPI 255


Search completed: October 29, 2003, 18:36:19
Job time : 4.19149 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein protein search, using sw model

Run on: October 29, 2003, 18:32:09 ; Search time 1.73308 Seconds
(without alignments)
3107.451 Million cell updates/sec

Title: US-09-980-043-4
Perfect score: 56
Sequence: 1 MMESGAPICHTCGEQVGHDA... .....KSCFEFEIKEGRKVCLRCGS 56

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283308 seqs, 96168682 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR_76:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	8	14.3	326	2	T08591	TGACG-motif bindin
2	8	14.3	958	2	T04870	cellulose synthase
3	7	12.5	44	2	G64536	hypothetical prote
4	7	12.5	154	2	T25602	hypothetical prote
5	7	12.5	312	2	E71624	rifin PFB0055c - m
6	7	12.5	322	2	T08592	TGACG-motif-bindin
7	7	12.5	322	2	T12093	TGACG-motif bindin
8	7	12.5	974	2	T10797	cellulose synthase
9	7	12.5	2895	2	H85362	hypothetical prote
10	6	10.7	91	2	E91173	HicA-like protein
11	6	10.7	91	2	E86019	HicA-like protein
12	6	10.7	102	2	T17741	hypothetical prote
13	6	10.7	112	2	D90214	conserved hypothet
14	6	10.7	128	2	G97176	stress-induced pro
15	6	10.7	152	2	S76079	hypothetical prote
16	6	10.7	159	2	S51498	dUTP diphosphatase
17	6	10.7	171	2	G84421	probable auxin-ind
18	6	10.7	176	2	F82478	hypothetical prote
19	6	10.7	202	2	S59955	hypothetical prote
20	6	10.7	219	2	B69463	hypothetical prote
21	6	10.7	229	2	AE1350	phosphoglyceromuta
22	6	10.7	229	2	E70471	conserved hypothet
23	6	10.7	229	2	S47859	GLC8 protein - yea
24	6	10.7	247	2	E71355	probable ribosomal
25	6	10.7	250	2	AC1147	probable conserved
26	6	10.7	250	2	AC1506	probable conserved
27	6	10.7	253	2	AB3327	conserved hypothet
28	6	10.7	256	2	AI1204	molybdate ABC tran
29	6	10.7	256	2	D86874	ABC transporter AT
30	6	10.7	257	2	AH1561	molybdate ABC tran
31	6	10.7	258	1	F69374	conserved hypothet
32	6	10.7	261	2	S09647	gentamycin 3'-N-ac
33	6	10.7	289	2	T03462	conserved hypothet
34	6	10.7	330	2	D71600	rifin PFB1035w - m
35	6	10.7	342	2	AF2823	hypothetical prote
36	6	10.7	342	2	E97601	agta [imported] -
37	6	10.7	349	2	A88197	protein ZK1127.4 [
38	6	10.7	359	2	S35157	Delta6 fatty acid
39	6	10.7	366	2	AE2848	conserved hypothet
40	6	10.7	366	2	C97625	hypothetical prote
41	6	10.7	394	2	D75256	queuine tRNA ribos
42	6	10.7	395	1	WZECCB	cystathionine beta
43	6	10.7	395	2	JV0020	cystathionine beta
44	6	10.7	395	2	D91115	cystathionine beta
45	6	10.7	395	2	D85960	cystathionine beta

ALIGNMENTS

RESULT 1
T08591
TGACG-motif binding protein STF1 - soybean
C;Species: Glycine max (soybean)
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 11-Jun-1999
C;Accession: T08591
R;Cheong, Y.H.; Yoo, C.M.; Park, J.M.; Ryu, G.R.; Goekjian, V.H.; Nagao, R.T.; Key, J
submitted to the EMBL Data Library, September 1995
A;Description: STF1 is a novel TGACG-binding factor with a zinc-finger motif and a bZ
A;Reference number: Z16445
A;Accession: T08591
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-326 <CHE>
A;Cross-references: EMBL:L28003; NID:g2934883; PID:g2934884
A;Experimental source: strain Williams; hypocotyl

Query Match 14.3%; Score 8; DB 2; Length 326;
Best Local Similarity 100.0%; Pred. No. 0.87;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 24 LFVACHEC 31
||||||||
Db 54 LFVACHEC 61

RESULT 2
T04870
cellulose synthase (EC 2.4.1.-) catalytic chain F28A21.190 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-Jun-1999
C;Accession: T04870
R;Bevan, M.; Mueller, M.W.; Muendlein, A.; Felber, R.; Bancroft, I.; Mewes, H.W.; May
submitted to the Protein Sequence Database, February 1999
A;Reference number: Z15387
A;Accession: T04870
A;Molecule type: DNA
A;Residues: 1-958 <BEV>
A;Cross-references: EMBL:AL035526
A;Experimental source: cultivar Columbia; BAC clone F28A21
C;Genetics:
A;Map position: 4
A;Introns: 60/1; 76/3; 94/1; 149/2; 240/2; 355/3; 401/3; 443/3; 514/3; 662/3; 780/3;
A;Note: F28A21.190
C;Keywords: glycosyltransferase; hexosyltransferase

Query Match 14.3%; Score 8; DB 2; Length 958;
Best Local Similarity 100.0%; Pred. No. 2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 25 FVACHECS 32
|||||||

```
o          25 FVACHECS 32


ESULT 3
 4536
ypothetical protein (HP0135, jhp0123) - Helicobacter pylori
 Species: Helicobacter pylori
 Variety: strains J99, 26695
 Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 08-Oct-1999
 Accession: G64536; E71971
 Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.
 eterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenne
 n, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.
 ture 388, 539-547, 1997
 Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.
 Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
 Reference number: A64520; MUID:97394467; PMID:9252185
 Accession: G64536
 Status: nucleic acid sequence not shown; translation not shown
 Molecule type: DNA
 Residues: 1-44 <TOM>
 Cross-references: GB:AE000535; GB:AE000511; NID:g2313217; PIDN:AAD07212.1; PID:g231322
 Experimental source: strain 26695
 Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;
 Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;
 ature 397, 176-180, 1999
 Title: Genomic sequence comparison of two unrelated isolates of the human gastric path
 Reference number: A71800; MUID:99120557; PMID:9923682
 Accession: E71971
 Molecule type: DNA
 Residues: 1-44 <ARN>
 Cross-references: GB:AE001451; GB:AE001439; NID:g4154626; PIDN:AAD05700.1; PID:g415463
 Experimental source: strain J99
 Genetics:
 Gene: HP0135; jhp0123

 Query Match             12.5%;  Score 7;  DB 2;  Length 44;
 Best Local Similarity   100.0%;  Pred. No. 1.9;
 Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

y          24 LFVACHE 30
           |||||||
o          13 LFVACHE 19


ESULT 4
25602
ypothetical protein C32E8.1 - Caenorhabditis elegans
 Species: Caenorhabditis elegans
 Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 Accession: T25602
 Gattung, S.
ubmitted to the EMBL Data Library, February 1997
 Description: The sequence of C. elegans cosmid C32E8.
 Reference number: Z20056
 Accession: T25602
 Status: preliminary; translated from GB/EMBL/DDBJ
 Molecule type: DNA
 Residues: 1-154 <GAT>
 Cross-references: EMBL:U88308; PIDN:AAB42321.1; GSPDB:GN00019; CESP:C32E8.1
,Experimental source: strain Bristol N2; clone C32E8
,Genetics:
,Gene: CESP:C32E8.1
,Map position: 1
,Introns: 107/1

 Query Match             12.5%;  Score 7;  DB 2;  Length 154;
 Best Local Similarity   100.0%;  Pred. No. 5.3;
 Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

y          14 EQVGHDA 20
           |||||||
o          13 EQVGHDA 19


RESULT 5
E71624
rifin PFB0055c - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 02-Mar-2001
C;Accession: E71624
R;Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
.; Pertea, M.; Salzberg, S.; Mau, B.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.
Science 282, 1126-1132, 1998
A;Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A;Reference number: A71600; MUID:99021743; PMID:9804551
A;Accession: E71624
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-312 <GAR>
A;Cross-references: GB:AE001375; GB:AE001375; NID:g1845879; PIDN:AAC71801.1; PID:g384508
A;Experimental source: clone 3D7
C;Genetics:
A;Gene: PFB0055c
C;Superfamily: Plasmodium falciparum rifin PFF1005w

  Query Match             12.5%;  Score 7;  DB 2;  Length 312;
  Best Local Similarity   100.0%;  Pred. No. 9.2;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         49 KVCLRCG 55
              |||||||
Db        113 KVCLRCG 119


RESULT 6
T08592
TGACG-motif-binding protein STF2 - soybean
C;Species: Glycine max (soybean)
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 11-Jun-1999
C;Accession: T08592
R;Cheong, Y.H.; Yoo, C.M.; Park, J.M.; Ryu, G.R.; Goekjian, V.H.; Nagao, R.T.; Key, J.L.
submitted to the EMBL Data Library, September 1995
A;Description: STF1 is a novel TGACG-binding factor with a zinc-finger motif and a bZIP
A;Reference number: Z16445
A;Accession: T08592
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-322 <CHE>
A;Cross-references: EMBL:L28004; NID:g986966; PID:g2934885
A;Experimental source: strain Williams; hypocotyl
C;Genetics:
A;Gene: STF2

  Query Match             12.5%;  Score 7;  DB 2;  Length 322;
  Best Local Similarity   100.0%;  Pred. No. 9.4;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         25 FVACHEC 31
              |||||||
Db         55 FVACHEC 61


RESULT 7
T12093
TGACG-motif binding protein - fava bean
C;Species: Vicia faba (fava bean)
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 20-Jun-2000
C;Accession: T12093
R;Wohlfarth, T.
submitted to the EMBL Data Library, May 1996
A;Reference number: Z17415
A;Accession: T12093
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-322 <WOH>
```

A;Cross-references: EMBL:X97904

Query Match 12.5%; Score 7; DB 2; Length 322;
Best Local Similarity 100.0%; Pred. No. 9.4;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 25 FVACHEC 31
|||||||
Db 52 FVACHEC 58

RESULT 8
T10797
cellulose synthase (EC 2.4.1.-) catalytic chain celA1 - upland cotton
C;Species: Gossypium hirsutum (upland cotton)
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
C;Accession: T10797
R;Pear, J.R.; Kawagoe, Y.; Schreckengost, W.E.; Delmer, D.P.; Stalker, D.M.
Proc. Natl. Acad. Sci. U.S.A. 93, 12637-12642, 1996
A;Title: Higher plants contain homologs of the bacterial celA genes encoding the catalyt
A;Reference number: Z17152; MUID:97057296; PMID:8901635
A;Accession: T10797
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-974 <PEA>
A;Cross-references: EMBL:U58283; NID:g1706955; PIDN:AAB67766.1; PID:g1706956
A;Experimental source: strain Acala SJ-2; fiber
C;Genetics:
A;Gene: celA1
C;Function:
A;Description: involved in the synthesis of cellulose
C;Keywords: glycosyltransferase; hexosyltransferase

Query Match 12.5%; Score 7; DB 2; Length 974;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 25 FVACHEC 31
|||||||
Db 25 FVACHEC 31

RESULT 9
H85362
hypothetical protein AT4g30990 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 16-Feb-2001
C;Accession: H85362
R;Anonymous; The European Union Arabidopsis Genome Sequencing Consortium, The Cold Sprin
Nature 402, 769-777, 1999
A;Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
A;Reference number: A85001; MUID:20084484; PMID:10617
A;Accession: H85362
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-2895 <STO>
A;Cross-references: GB:NC_001268; NID:g7270001; PIDN:CAB79817.1; GSPDB:GN00140
C;Genetics:
A;Gene: AT4g30990
A;Map position: 4

Query Match 12.5%; Score 7; DB 2; Length 2895;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 44 IKEGRKV 50
|||||||
Db 275 IKEGRKV 281

RESULT 10
E91173
HicA-like protein [imported] - Escherichia coli (strain O157:H7, substrain RIMD 0509952)
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C;Accession: E91173
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and g
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: E91173
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-91 <HAY>
A;Cross-references: GB:BA000007; PIDN:BAB37780.1; PID:g13363831; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs4357

Query Match 10.7%; Score 6; DB 2; Length 91;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 43 EIKEGR 48
||||||
Db 43 EIKEGR 48

RESULT 11
E86019
HicA-like protein [imported] - Escherichia coli (strain O157:H7, substrain EDL933)
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Nov-2001
C;Accession: E86019
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayh
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodac
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: E86019
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-91 <STO>
A;Cross-references: GB:AE005174; NID:g12518173; PIDN:AAG58617.1; GSPDB:GN00145; UWGP:
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: Z4883

Query Match 10.7%; Score 6; DB 2; Length 91;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 43 EIKEGR 48
||||||
Db 43 EIKEGR 48

RESULT 12
T17741
hypothetical protein a249L - Chlorella virus PBCV-1
C;Species: Chlorella virus PBCV-1
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T17741
R;Graves, M.V.; Van Etten, J.L.
submitted to the EMBL Data Library, May 1999
A;Reference number: Z18806
A;Accession: T17741
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-102 <GRA>
A;Cross-references: EMBL:U42580; NID:g4028896; PIDN:AAC96617.1
A;Experimental source: specific host Chlorella strain NC64A
C;Genetics:
A;Note: a249L

```
Query Match             10.7%;  Score 6;  DB 2;  Length 102;
Best Local Similarity   100.0%;  Pred. No. 42;
Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

/         22 GELFVA 27
             ||||||
>         20 GELFVA 25


ESULT 13
90214
onserved hypothetical protein [imported] - Sulfolobus solfataricus
 Species: Sulfolobus solfataricus
 Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 24-May-2001
 Accession: D90214
 She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, P
rett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.
ibmitted to GenBank, April 2001
 Description: Sulfolobus solfataricus complete genome.
 Reference number: A99139
 Accession: D90214
 Status: preliminary
 Molecule type: DNA
 Residues: 1-112 <KUR>
 Cross-references: GB:AE006641; NID:g13813835; PIDN:AAK40971.1; GSPDB:GN00155
 Genetics:
 Gene: SSO0668

 Query Match             10.7%;  Score 6;  DB 2;  Length 112;
 Best Local Similarity   100.0%;  Pred. No. 45;
 Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

/         45 KEGRKV 50
             ||||||
>         94 KEGRKV 99


SSULT 14
97176
.ress-induced protein OsmC [imported] - Clostridium acetobutylicum
 Species: Clostridium acetobutylicum
 Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
 Accession: G97176
 Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
  Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
  Bacteriol. 183, 4823-4838, 2001
 Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo
 Reference number: A96900; MUID:21359325; PMID:21359325
 Accession: G97176
 Status: preliminary
 Molecule type: DNA
 Residues: 1-128 <KUR>
 Cross-references: GB:AE001437; PIDN:AAK80202.1; PID:g15025245; GSPDB:GN00168
 Experimental source: Clostridium acetobutylicum ATCC824
 Genetics:
 Gene: CAC2245

 Query Match             10.7%;  Score 6;  DB 2;  Length 128;
 Best Local Similarity   100.0%;  Pred. No. 50;
 Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

/          41 EFEIKE 46
              ||||||
>         121 EFEIKE 126


ESULT 15
76079
/pothetical protein sll0172 - Synechocystis sp. (strain PCC 6803)
 Species: Synechocystis sp.
 Variety: PCC 6803
 Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
 C;Accession: S76079
 R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
 o, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
 DNA Res. 3, 109-136, 1996
 A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis
 s.
 A;Reference number: S74322; MUID:97061201; PMID:8905231
 A;Accession: S76079
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-152 <KAN>
 A;Cross-references: EMBL:D63999; GB:AB001339; NID:g1001396; PIDN:BAA10057.1; PID:g100143
 A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
 C;Superfamily: Synechocystis hypothetical protein sll0172

  Query Match             [illegible]  Length 152;
  Best Local Similarity   [illegible]
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 ESGAPI 8
              ||||||
Db         94 ESGAPI 99


Search completed: October 29, 2003, 18:16:59
Job time : 4.73308 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 31, 2003, 23:55:54 ; Search time 42.7853 Seconds
(without alignments)
3559.830 Million cell updates/sec

Title: US-09-980-043-4
Perfect score: 56
Sequence: 1 MMESGAPICHTCGEQVGHDA.....KSCFEFEIKEGRKVCLRCGS 56

Scoring table: OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1811591 seqs, 1359896290 residues

Word size: 1

Total number of hits satisfying chosen parameters: 3617484

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=x1h
-Q=/cgn2_1/USPTO_spool/US09980043/runat_29102003_163932_22798/app_query.fasta_1.1358
-DB=Published_Applications_NA -QFMT=fastap -SUFFIX=oligo.rnpb -MINMATCH=0.1
-LOCPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=oligo
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1
-ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09980043 @CGN_1_1_350_@runat_29102003_163932_22798
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	7	12.5	133	10	US-09-878-574-10070	Sequence 10070, A
	2	7	12.5	235	10	US-09-881-752A-169	Sequence 169, App
	3	7	12.5	379	12	US-09-814-353-17542	Sequence 17542, A
	4	7	12.5	467	10	US-09-974-300-7577	Sequence 7577, Ap
c	5	7	12.5	585	13	US-10-027-632-230374	Sequence 230374,
	6	7	12.5	651	10	US-09-893-737-289	Sequence 289, App
c	7	7	12.5	744	13	US-10-027-632-103853	Sequence 103853,
c	8	7	12.5	744	13	US-10-027-632-323569	Sequence 323569,
	9	7	12.5	848	10	US-09-974-300-2915	Sequence 2915, Ap
c	10	7	12.5	961	13	US-10-027-632-171512	Sequence 171512,
	11	7	12.5	975	13	US-10-027-632-323555	Sequence 323555,
c	12	7	12.5	1203	14	US-10-156-761-5768	Sequence 5768, Ap
c	13	7	12.5	1292	13	US-10-027-632-251762	Sequence 251762,
c	14	7	12.5	1870	13	US-10-027-632-255980	Sequence 255980,
c	15	7	12.5	1870	13	US-10-027-632-255981	Sequence 255981,
c	16	7	12.5	1870	13	US-10-027-632-255982	Sequence 255982,
c	17	7	12.5	1870	13	US-10-027-632-255983	Sequence 255983,
c	18	7	12.5	1870	13	US-10-027-632-255984	Sequence 255984,
c	19	7	12.5	1914	14	US-10-205-841-13	Sequence 13, Appl
	20	7	12.5	2015	13	US-10-027-632-97256	Sequence 97256, A
	21	7	12.5	2799	14	US-10-161-803-33	Sequence 33, Appl
	22	7	12.5	3328	10	US-09-838-539-1	Sequence 1, Appli
	23	7	12.5	6510	12	US-10-311-455-367	Sequence 367, App
	24	7	12.5	8688	10	US-09-938-842A-1748	Sequence 1748, Ap
	25	7	12.5	14924	12	US-10-311-455-198	Sequence 198, App
	26	7	12.5	14924	12	US-10-240-452-22	Sequence 22, Appl
c	27	7	12.5	66109	10	US-09-880-107-3768	Sequence 3768, Ap
c	28	7	12.5	66109	12	US-09-873-367C-1060	Sequence 1060, Ap
c	29	7	12.5	198285	10	US-09-880-107-3814	Sequence 3814, Ap
c	30	7	12.5	251364	14	US-10-175-523-58	Sequence 58, Appl
c	31	7	12.5	251364	14	US-10-175-523-61	Sequence 61, Appl
c	32	7	12.5	251364	14	US-10-175-523-79	Sequence 79, Appl
	33	7	12.5	3673778	12	US-10-312-841-1	Sequence 1, Appli
	34	7	12.5	9025608	14	US-10-156-761-1	Sequence 1, Appli
	35	6	10.7	25	10	US-09-895-141-28	Sequence 28, Appl
c	36	6	10.7	25	14	US-10-098-263B-116428	Sequence 116428,
c	37	6	10.7	35	10	US-09-880-887-16	Sequence 16, Appl
c	38	6	10.7	60	12	US-09-908-975-12616	Sequence 12616, A
c	39	6	10.7	65	12	US-09-908-975-2988	Sequence 2988, Ap
c	40	6	10.7	65	12	US-10-032-585-117	Sequence 117, App
c	41	6	10.7	76	9	US-09-864-761-25825	Sequence 25825, A
	42	6	10.7	108	9	US-09-864-761-29954	Sequence 29954, A
	43	6	10.7	115	10	US-09-969-373-272	Sequence 272, App
	44	6	10.7	118	9	US-09-563-817-491	Sequence 491, App
	45	6	10.7	126	10	US-09-867-701-5501	Sequence 5501, Ap

ALIGNMENTS

RESULT 1
US-09-878-574-10070
; Sequence 10070, Application US/09878574
; Patent No. US20020110548A1
; GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Thompson, Michael D.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(15401)B
; CURRENT APPLICATION NUMBER: US/09/878,574
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/333,535
; PRIOR FILING DATE: 1999-06-14
; NUMBER OF SEQ ID NOS: 15775
; SEQ ID NO 10070
; LENGTH: 133
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: 701103248H1
US-09-878-574-10070

Alignment Scores:
Pred. No.: 14.6 Length: 133

```
core:                 7.00          Matches:       7
ercent Similarity:    100.00%       Conservative:  0
est Local Similarity: 100.00%       Mismatches:    0
uery Match:           12.50%        Indels:        0
B:                    10            Gaps:          0

S-09-980-043-4 (1-56) x US-09-878-574-10070 (1-133)

y         43 GluIleLysGluGlyArgLys 49
             |||||||||||||||||||||
b         36 GAGATCAAGGAAGGAAGGAAG 56

ESULT 2
S-09-881-752A-169
 Sequence 169, Application US/09881752A
 Patent No. US20020115078A1
 GENERAL INFORMATION:
  APPLICANT: Kleanthous, Harold
  APPLICANT:  Al-Garawi, Amal
  APPLICANT:  Miller, Charles
  APPLICANT:  Tomb, Jean-Francois
  APPLICANT:  Oomen, Raymond P.
  TITLE OF INVENTION: Identification of Polynucleotides
  TITLE OF INVENTION:  Encoding No. US20020115078A1el Helicobacter Polypeptides in the
  TITLE OF INVENTION:  Genome
  FILE REFERENCE: 06132/041002
  CURRENT APPLICATION NUMBER: US/09/881,752A
  CURRENT FILING DATE:  2001-06-15
  PRIOR APPLICATION NUMBER: US 08/833,457
  PRIOR FILING DATE: 1997-04-01
  NUMBER OF SEQ ID NOS: 370
  SOFTWARE: FastSEQ for Windows Version 4.0
 SEQ ID NO 169
   LENGTH: 235
   TYPE: DNA
   ORGANISM: Helicobacter pylori
   FEATURE:
   NAME/KEY: CDS
   LOCATION: (51)...(182)
JS-09-881-752A-169

\lignment Scores:
Pred. No.:            24.1          Length:        235
Score:                7.00          Matches:       7
Percent Similarity:   100.00%       Conservative:  0
Best Local Similarity: 100.00%      Mismatches:    0
Query Match:          12.50%        Indels:        0
DB:                   10            Gaps:          0

US-09-980-043-4 (1-56) x US-09-881-752A-169 (1-235)

Qy         24 LeuPheValAlaCysHisGlu 30
              ||||||||||||||||||||
Db         87 TTGTTTGTGGCTTGCCACGAA 107

RESULT 3
US-09-814-353-17542
; Sequence 17542, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
;  APPLICANT: Lee, John
;  APPLICANT:  Thompson, Pamela
;  APPLICANT:  Lillie, James
;  TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
;  TITLE OF INVENTION:  IDENTIFICATION, ASSESSMENT, PREVENTION, AND
;  TITLE OF INVENTION:  THERAPY OF OVARIAN CANCER
;  FILE REFERENCE: MRI-006B
;  CURRENT APPLICATION NUMBER: US/09/814,353
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: US 60/191,031
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: US 60/207,124
;  PRIOR FILING DATE: 2000-05-25
;  PRIOR APPLICATION NUMBER: US 60/211,940
;  PRIOR FILING DATE: 2000-06-15
;  PRIOR APPLICATION NUMBER: US 60/216,820
;  PRIOR FILING DATE: 2000-07-07
;  PRIOR APPLICATION NUMBER: US 60/220,661
;  PRIOR FILING DATE: 2000-07-25
;  PRIOR APPLICATION NUMBER: US 60/257,672
;  PRIOR FILING DATE: 2000-12-21
;  NUMBER OF SEQ ID NOS: 22047
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17542
;   LENGTH: 379
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-814-353-17542

Alignment Scores:
Pred. No.:            [illegible]   Length:        379
Score:                7.00          Matches:       7
Percent Similarity:   100.00%       Conservative:  0
Best Local Similarity: 100.00%      Mismatches:    0
Query Match:          12.50%        Indels:        0
DB:                   12            Gaps:          0

US-09-980-043-4 (1-56) x US-09-814-353-17542 (1-379)

Qy         49 LysValCysLeuArgCysGly 55
              |||||||||||||||:|||||
Db         81 AAGGTATGCTTGAGGTGTGGC 101

RESULT 4
US-09-974-300-7577
; Sequence 7577, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
;  APPLICANT: Berka, Randy M.
;  APPLICANT:  Clausen, Ib Groth
;  TITLE OF INVENTION: Methods For Monitoring Multiple Gene
;  TITLE OF INVENTION:  Expression
;  FILE REFERENCE: 10085.500-US
;  CURRENT APPLICATION NUMBER: US/09/974,300
;  CURRENT FILING DATE:  2001-10-05
;  PRIOR APPLICATION NUMBER: 09/680,598
;  PRIOR FILING DATE: 2000-10-06
;  PRIOR APPLICATION NUMBER: 60/279,526
;  PRIOR FILING DATE: 2001-03-27
;  NUMBER OF SEQ ID NOS: 8481
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7577
;   LENGTH: 467
;   TYPE: DNA
;   ORGANISM: Bacillus clausii
US-09-974-300-7577

Alignment Scores:
Pred. No.:            44.1          Length:        467
Score:                7.00          Matches:       7
Percent Similarity:   100.00%       Conservative:  0
Best Local Similarity: 100.00%      Mismatches:    0
Query Match:          12.50%        Indels:        0
DB:                   10            Gaps:          0

US-09-980-043-4 (1-56) x US-09-974-300-7577 (1-467)

Qy         44 IleLysGluGlyArgLysVal 50
              |||||||||||||||||||||
Db        397 ATAAAGGAAGGCCGGAAAGTC 417

RESULT 5
US-10-027-632-230374/c
; Sequence 230374, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
```

```
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 230374
;   LENGTH: 585
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-230374

Alignment Scores:
Pred. No.:                53.7          Length:          585
Score:                    7.00          Matches:         7
Percent Similarity:       100.00%       Conservative:    0
Best Local Similarity:    100.00%       Mismatches:      0
Query Match:              12.50%        Indels:          0
DB:                       13            Gaps:            0

US-09-980-043-4 (1-56) x US-10-027-632-230374 (1-585)

Qy          43 GluIleLysGluGlyArgLys 49
               |||||||||||||||||||
Db         558 GAAATAAAAGAAGGAAGGAAA 538

RESULT 6
US-09-893-737-289
; Sequence 289, Application US/09893737
; Patent No. US20020110855A1
; GENERAL INFORMATION:
;  APPLICANT: Sheppard, Paul O.
;  APPLICANT:  Presnell, Scott R.
;  TITLE OF INVENTION: MAMMALIAN SECRETED PROTEINS
;  FILE REFERENCE: 00-41
;  CURRENT APPLICATION NUMBER: US/09/893,737
;  CURRENT FILING DATE:  2001-06-28
;  PRIOR APPLICATION NUMBER: US 60/215,449
;  PRIOR FILING DATE: 2000-06-30
;  NUMBER OF SEQ ID NOS: 329
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 289
;   LENGTH: 651
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(651)
US-09-893-737-289

Alignment Scores:
Pred. No.:                59            Length:          651
Score:                    7.00          Matches:         7
Percent Similarity:       100.00%       Conservative:    0
Best Local Similarity:    100.00%       Mismatches:      0
Query Match:              12.50%        Indels:          0
DB:                       10            Gaps:            0

US-09-980-043-4 (1-56) x US-09-893-737-289 (1-651)

Qy           3 GluSerGlyAlaProIleCys 9
               |||||||||||||||||||
Db         451 GAGAGTGGGGCACCCATCTGC 471

RESULT 7
US-10-027-632-103853/c
; Sequence 103853, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 103853
;   LENGTH: 744
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-103853

Alignment Scores:
Pred. No.:                66.4          Length:          744
Score:                    7.00          Matches:         7
Percent Similarity:       100.00%       Conservative:    0
Best Local Similarity:    100.00%       Mismatches:      0
Query Match:              12.50%        Indels:          0
DB:                       13            Gaps:            0

US-09-980-043-4 (1-56) x US-10-027-632-103853 (1-744)

Qy          34 ProMetCysLysSerCysPhe 40
               |||||||||||||||||||
Db          51 CCTATGTGTAAGAGTTGTTTC 31

RESULT 8
US-10-027-632-323569/c
; Sequence 323569, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
```

PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 323569
LENGTH: 744
TYPE: DNA
ORGANISM: Human
S-10-027-632-323569

ignment Scores:

red. No.:	66.4	Length:	744
core:	7.00	Matches:	7
ercent Similarity:	100.00%	Conservative:	0
est Local Similarity:	100.00%	Mismatches:	0
uery Match:	12.50%	Indels:	0
B:	13	Gaps:	0

S-09-980-043-4 (1-56) x US-10-027-632-323569 (1-744)

```
y          34 ProMetCysLysSerCysPhe 40
              ||||||||||||||||||||
b          51 CCTATGTGTAAGAGTTGTTTC 31
```

ESULT 9
S-09-974-300-2915
Sequence 2915, Application US/09974300
Patent No. US20020146721A1
GENERAL INFORMATION:
APPLICANT: Berka, Randy M.
APPLICANT: Clausen, Ib Groth
TITLE OF INVENTION: Methods For Monitoring Multiple Gene
TITLE OF INVENTION: Expression
FILE REFERENCE: 10085.500-US
CURRENT APPLICATION NUMBER: US/09/974,300
CURRENT FILING DATE: 2001-10-05
PRIOR APPLICATION NUMBER: 09/680,598
PRIOR FILING DATE: 2000-10-06
PRIOR APPLICATION NUMBER: 60/279,526
PRIOR FILING DATE: 2001-03-27
NUMBER OF SEQ ID NOS: 8481
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 2915
LENGTH: 848
TYPE: DNA
ORGANISM: Bacillus licheniformis
S-09-974-300-2915

lignment Scores:

red. No.:	74.4	Length:	848
core:	7.00	Matches:	7
ercent Similarity:	100.00%	Conservative:	0
est Local Similarity:	100.00%	Mismatches:	0
uery Match:	12.50%	Indels:	0
B:	10	Gaps:	0

S-09-980-043-4 (1-56) x US-09-974-300-2915 (1-848)

```
y          50 ValCysLeuArgCysGlySer 56
              ||||||||||||||||||||
b         740 GTTTGTTTAAGATGTGGGAGC 760
```

ESULT 10
S-10-027-632-171512/c
Sequence 171512, Application US/10027632
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
TITLE OF INVENTION: Polymorphisms in the Human Genome
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30

; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-0[illegible]12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-0[illegible]29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 171512
; LENGTH: 961
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-171512

Alignment Scores:

Pred. No.:	83.2	Length:	961
Score:	7.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	12.50%	Indels:	0
DB:	13	Gaps:	0

US-09-980-043-4 (1-56) x US-10-027-632-171512 (1-961)

```
Qy         45 LysGluGlyArgLysValCys 61
              |||||||||||||||||||
Db        506 AAGGAAGGAAGGAAAGTGTGT 486
```

RESULT 11
US-10-027-632-323555
; Sequence 323555, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-04-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 323555
; LENGTH: 975
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-323555

Alignment Scores:

Pred. No.:	84.1	Length:	975
Score:	7.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0

Query Match: 12.50% Indels: 0
DB: 13 Gaps: 0

US-09-980-043-4 (1-56) x US-10-027-632-323555 (1-975)

Qy 34 ProMetCysLysSerCysPhe 43
||||||||||||||||||||
Db 693 CCTATGTGTAAGAGTTGTTTC 713

RESULT 12
US-10-156-761-5768/c
; Sequence 5768, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 5768
; LENGTH: 1203
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1203)
US-10-156-761-5768

Alignment Scores:
Pred. No.: 101 Length: 1203
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 12.50% Indels: 0
DB: 14 Gaps: 0

US-09-980-043-4 (1-56) x US-10-156-761-5768 (1-1203)

Qy 13 GlyGluGlnValGlyHisAsp 19
||||||||||||||||||||
Db 816 GGCGAGCAGGTCGGGCACGAC 836

RESULT 13
US-10-027-632-251762/c
; Sequence 251762, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 251762
; LENGTH: 1292
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-251762

Alignment Scores:
Pred. No.: 108 Length: 1292
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 12.50% Indels: 0
DB: 13 Gaps: 0

US-09-980-043-4 (1-56) x US-10-027-632-251762 (1-1292)

Qy 47 GlyArgLysValCysLeuArg 53
||||||||||||||||||||
Db 118 GGAAGAAAAGTCTGTTTACGT 98

RESULT 14
US-10-027-632-255980/c
; Sequence 255980, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 255980
; LENGTH: 1870
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-255980

Alignment Scores:
Pred. No.: 149 Length: 1870
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 12.50% Indels: 0
DB: 13 Gaps: 0

US-09-980-043-4 (1-56) x US-10-027-632-255980 (1-1870)

Qy 43 GluIleLysGluGlyArgLys 49
||||||||||||||||||||
Db 1422 GAGATAAAGGAAGGGAGGAAA 1402

RESULT 15

```
S-10-027-632-255981/c
Sequence 255981, Application US/10027632
GENERAL INFORMATION:
 APPLICANT: Wang, David G.
 TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
 TITLE OF INVENTION:  Polymorphisms in the Human Genome
 FILE REFERENCE: 108827.129
 CURRENT APPLICATION NUMBER: US/10/027,632
 CURRENT FILING DATE:  2002-04-30
 PRIOR APPLICATION NUMBER: US 60/218,006
 PRIOR FILING DATE: 2000-07-12
 PRIOR APPLICATION NUMBER: US 60/198,676
 PRIOR FILING DATE: 2000-04-20
 PRIOR APPLICATION NUMBER: US 60/193,483
 PRIOR FILING DATE: 2000-03-29
 PRIOR APPLICATION NUMBER: US 60/185,218
 PRIOR FILING DATE: 2000-02-24
 PRIOR APPLICATION NUMBER: US 60/167,363
 PRIOR FILING DATE: 1999-11-23
 PRIOR APPLICATION NUMBER: US 60/156,358
 PRIOR FILING DATE: 1999-09-28
 PRIOR APPLICATION NUMBER: US 60/146,002
 PRIOR FILING DATE: 1999-08-09
 NUMBER OF SEQ ID NOS: 325720
 SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 255981
  LENGTH: 1870
  TYPE: DNA
  ORGANISM: Human
-10-027-632-255981

ignment Scores:
ed. No.:                149             Length:         1870
ore:                    7.00            Matches:        7
rcent Similarity:       100.00%         Conservative:   0
st Local Similarity:    100.00%         Mismatches:     0
ery Match:              12.50%          Indels:         0
:                       13              Gaps:           0

-09-980-043-4 (1-56) x US-10-027-632-255981 (1-1870)

          43 GluIleLysGluGlyArgLys 49
             ||||||||||||||||||||
        1422 GAGATAAAGGAAGGGAGGAAA 1402


arch completed: November  1, 2003, 07:29:50
b time : 50.7853 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 31, 2003, 23:31:14 ; Search time 8.93617 Seconds
(without alignments)
2766.001 Million cell updates/sec

Title: US-09-980-043-4
Perfect score: 56
Sequence: 1 MMESGAPICHTCGEQVGHDA.....KSCFEFEIKEGRKVCLRCGS 56

Scoring table: OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 569978 seqs, 220691566 residues

Word size: 1

Total number of hits satisfying chosen parameters: 1135288

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09980043/runat_29102003_163931_22725/app_query.fasta_1.1358
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=oligo.rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL
OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09980043_@CGN_1_1_76_@runat_29102003_163931_22725 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	7	12.5	1461	4	US-09-328-352-2189	Sequence 2189, Ap
2	7	12.5	2232	4	US-09-328-352-947	Sequence 947, App
3	7	12.5	3328	3	US-08-960-048-1	Sequence 1, Appli
4	7	12.5	3328	4	US-09-838-586-1	Sequence 1, Appli
5	7	12.5	6678	4	US-09-268-347-33	Sequence 33, Appl
c 6	6	10.7	35	3	US-09-526-935B-16	Sequence 16, Appl
c 7	6	10.7	136	1	US-08-751-782-3	Sequence 3, Appli
c 8	6	10.7	136	2	US-08-925-171-3	Sequence 3, Appli
c 9	6	10.7	200	3	US-09-243-335-3	Sequence 3, Appli
c 10	6	10.7	238	2	US-08-332-766A-14	Sequence 14, Appl
c 11	6	10.7	253	2	US-08-520-678A-25	Sequence 25, Appl
c 12	6	10.7	253	3	US-08-897-126-25	Sequence 25, Appl
c 13	6	10.7	257	2	US-08-520-678A-24	Sequence 24, Appl
c 14	6	10.7	257	3	US-08-897-126-24	Sequence 24, Appl
15	6	10.7	261	4	US-09-313-294A-2372	Sequence 2372, Ap
16	6	10.7	271	4	US-09-313-294A-210	Sequence 210, App
c 17	6	10.7	285	2	US-08-520-678A-20	Sequence 20, Appl
c 18	6	10.7	285	3	US-08-897-126-20	Sequence 20, Appl
19	6	10.7	295	3	US-09-018-584A-8	Sequence 8, Appli
c 20	6	10.7	298	2	US-08-520-678A-21	Sequence 21, Appl
c 21	6	10.7	298	3	US-08-897-126-21	Sequence 21, Appl
c 22	6	10.7	310	4	US-08-979-847B-100	Sequence 100, App
c 23	6	10.7	321	2	US-08-520-678A-23	Sequence 23, Appl
c 24	6	10.7	321	3	US-08-897-126-23	Sequence 23, Appl
25	6	10.7	350	2	US-08-332-766A-20	Sequence 20, Appl
c 26	6	10.7	356	2	US-08-520-678A-22	Sequence 22, Appl
c 27	6	10.7	356	3	US-08-897-126-22	Sequence 22, Appl
c 28	6	10.7	388	2	US-08-623-906A-13	Sequence 13, Appl
c 29	6	10.7	440	4	US-09-118-554-32	Sequence 32, Appl
c 30	6	10.7	440	4	US-09-118-627-32	Sequence 32, Appl
c 31	6	10.7	440	4	US-09-602-877A-32	Sequence 32, Appl
32	6	10.7	441	4	US-09-328-352-3871	Sequence 3871, Ap
33	6	10.7	465	4	US-09-252-991A-3751	Sequence 3751, Ap
34	6	10.7	483	4	US-09-252-991A-5514	Sequence 5514, Ap
c 35	6	10.7	525	4	US-09-252-991A-5225	Sequence 5225, Ap
36	6	10.7	604	4	US-09-614-069-5	Sequence 5, Appli
c 37	6	10.7	700	4	US-09-222-938A-60	Sequence 60, Appl
c 38	6	10.7	753	3	US-09-232-191-14	Sequence 14, Appl
c 39	6	10.7	753	3	US-09-232-200-14	Sequence 14, Appl
c 40	6	10.7	753	4	US-09-232-197-14	Sequence 14, Appl
c 41	6	10.7	753	4	US-09-232-201-14	Sequence 14, Appl
42	6	10.7	757	3	US-08-998-416-579	Sequence 579, App
43	6	10.7	765	4	US-09-252-991A-14063	Sequence 14063, A
c 44	6	10.7	772	4	US-09-328-475C-134	Sequence 134, App
c 45	6	10.7	792	4	US-09-252-991A-5659	Sequence 5659, Ap

ALIGNMENTS

RESULT 1
US-09-328-352-2189
; Sequence 2189, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTE
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 2189
; LENGTH: 1461
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-2189

Alignment Scores:
Pred. No.: 53.2 Length: 1461
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 12.50% Indels: 0
DB: 4 Gaps: 0

US-09-980-043-4 (1-56) x US-09-328-352-2189 (1-1461)

Qy 36 CysLysSerCysPheGluPhe 42
||||||||||||||||||||||
Db 1365 TGCAAAAGTTGTTTCGAATTT 1385

RESULT 2
US-09-328-352-947
; Sequence 947, Application US/09328352
; Patent No. 6562958

```
GENERAL INFORMATION:
 APPLICANT: Gary L. Breton et al.
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
 TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
 FILE REFERENCE: GTC99-03PA
 CURRENT APPLICATION NUMBER: US/09/328,352
 CURRENT FILING DATE: 1999-06-04
 NUMBER OF SEQ ID NOS: 8252
SEQ ID NO 947
  LENGTH: 2232
  TYPE: DNA
  ORGANISM: Acinetobacter baumannii
3-09-328-352-947

ignment Scores:
ed. No.:                79.4          Length:        2232
core:                   7.00          Matches:       7
rcent Similarity:       100.00%       Conservative:  0
st Local Similarity:    100.00%       Mismatches:    0
ery Match:              12.50%        Indels:        0
:                       4             Gaps:          0

-09-980-043-4 (1-56) x US-09-328-352-947 (1-2232)

          9 CysHisThrCysGlyGluGln 15
            |||||||||||||||||||||
       2127 TGTCACACTTGCGGTGAACAA 2147

SULT 3
-08-960-048-1
Sequence 1, Application US/08960048C
Patent No. 6271443
GENERAL INFORMATION:
 APPLICANT: Stalker, D. et al.
 TITLE OF INVENTION: Plant Cellulose Synthase and Promoter
 TITLE OF INVENTION: Sequences
 FILE REFERENCE: 15621/01/US
 CURRENT APPLICATION NUMBER: US/08/960,048C
 CURRENT FILING DATE: 1997-10-29
 PRIOR APPLICATION NUMBER: 60/029,987
 PRIOR FILING DATE: 1996-10-29
 NUMBER OF SEQ ID NOS: 12
 SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 1
  LENGTH: 3328
  TYPE: DNA
  ORGANISM: Artificial Sequence
  FEATURE:
  OTHER INFORMATION: Synthetic Oligonucleotide
-08-960-048-1

ignment Scores:
ed. No.:                116           Length:        3328
ore:                    7.00          Matches:       7
rcent Similarity:       100.00%       Conservative:  0
st Local Similarity:    100.00%       Mismatches:    0
ery Match:              12.50%        Indels:        0
:                       3             Gaps:          0

-09-980-043-4 (1-56) x US-08-960-048-1 (1-3328)

         25 PheValAlaCysHisGluCys 31
            |||||||||||||||||||||
        251 TTTGTGGCTTGCCATGAATGT 271

SULT 4
-09-838-586-1
Sequence 1, Application US/09838586
Patent No. 6576818
GENERAL INFORMATION:
 APPLICANT: Stalker, D. et al.
 TITLE OF INVENTION: Plant Cellulose Synthase and Promoter
 TITLE OF INVENTION: Sequences
; FILE REFERENCE: 15621/02/US
; CURRENT APPLICATION NUMBER: US/09/838,586
; CURRENT FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/029,987
; PRIOR FILING DATE: 1996-10-29
; PRIOR APPLICATION NUMBER: 08/960,048
; PRIOR FILING DATE: 1997-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 3328
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic Oligonucleotide
US-09-838-586-1

Alignment Scores:
Pred. No.:                116           Length:        3328
Score:                    7.00          Matches:       7
Percent Similarity:       100.00%       Conservative:  0
Best Local Similarity:    100.00%       Mismatches:    0
Query Match:              12.50%        Indels:        0
DB:                       4             Gaps:          0

US-09-980-043-4 (1-56) x US-09-838-586-1 (1-3328)

Qy          25 PheValAlaCysHisGluCys 31
               |||||||||||||||||||||
Db         251 TTTGTGGCTTGCCATGAATGT 271

RESULT 5
US-09-268-347-33
; Sequence 33, Application US/09268347
; Patent No. 6335182
; GENERAL INFORMATION:
;  APPLICANT: Loosmore, Sheena M.
;  TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS
;  FILE REFERENCE: 1038-860
;  CURRENT APPLICATION NUMBER: US/09/268,347
;  CURRENT FILING DATE: 1999-03-16
;  NUMBER OF SEQ ID NOS: 54
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 33
;   LENGTH: 6678
;   TYPE: DNA
;   ORGANISM: Haemophilus influenzae
US-09-268-347-33

Alignment Scores:
Pred. No.:                223           Length:        6678
Score:                    7.00          Matches:       7
Percent Similarity:       100.00%       Conservative:  0
Best Local Similarity:    100.00%       Mismatches:    0
Query Match:              12.50%        Indels:        0
DB:                       4             Gaps:          0

US-09-980-043-4 (1-56) x US-09-268-347-33 (1-6678)

Qy          38 SerCysPheGluPheGluIle 44
               |||||||||||||||||||||
Db        3174 AGTTGTTTTGAGTTTGAAATT 3194

RESULT 6
US-09-526-935B-16/c
; Sequence 16, Application US/09526935B
; Patent No. 6271025
; GENERAL INFORMATION:
;  APPLICANT:  NEGRIER, CLAUDE
;  APPLICANT:  PLANTIER, JEAN LUC
;  TITLE OF INVENTION:  MODIFIED FACTOR VIII cDNA
;  FILE REFERENCE:  06478.1441
;  CURRENT APPLICATION NUMBER: US/09/526,935B
```

; CURRENT FILING DATE: 2000-03-16
; PRIOR APPLICATION NUMBER: EP 99104050.2
; PRIOR FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 16
; LENGTH: 35
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: oligonucleotide
US-09-526-935B-16

Alignment Scores:

Pred. No.:	21.1	Length:	35
Score:	6.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	10.71%	Indels:	0
DB:	3	Gaps:	0

US-09-980-043-4 (1-56) x US-09-526-935B-16 (1-35)

```
Qy         47 GlyArgLysValCysLeu 52
              ||||||||||||||||||
Db         24 GGACGCAAGGTTTGTTTA 7
```

RESULT 7
US-08-751-782-3/c
; Sequence 3, Application US/08751782
; Patent No. 5821352
; GENERAL INFORMATION:
; APPLICANT: Heintz, Nathaniel
; APPLICANT: Gubbay, Johnathan
; APPLICANT: Skinner, Michael
; TITLE OF INVENTION: A cDNA Library Prepared during
; TITLE OF INVENTION: Regression of Rat Prostate and Methods of Use Thereof
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David A. Jackson, Esq.
; STREET: 411 Hackensack Ave, Continental Plaza, 4th
; STREET: Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/751,782
; FILING DATE: 18-NOV-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 136 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; IMMEDIATE SOURCE:
; CLONE: 10.5
US-08-751-782-3

Alignment Scores:

Pred. No.:	76	Length:	136
Score:	6.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	10.71%	Indels:	0
DB:	1	Gaps:	0

US-09-980-043-4 (1-56) x US-08-751-782-3 (1-136)

```
Qy         14 GluGlnValGlyHisAsp 19
              ||||||||||||||||||
Db         23 GAGCAGGTAGGCCATGAT 6
```

RESULT 8
US-08-925-171-3/c
; Sequence 3, Application US/08925171
; Patent No. 5928871
; GENERAL INFORMATION:
; APPLICANT: Heintz, Nathaniel
; APPLICANT: Gubbay, Johnathan
; APPLICANT: Skinner, Michael
; TITLE OF INVENTION: A cDNA Library Prepared during
; TITLE OF INVENTION: Regression of Rat Prostate and Methods of Use Thereof
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David A. Jackson, Esq.
; STREET: 411 Hackensack Ave, Continental Plaza, 4th
; STREET: Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/925,171
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-190 CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 136 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; IMMEDIATE SOURCE:
; CLONE: 10.5
US-08-925-171-3

Alignment Scores:

Pred. No.:	76	Length:	136
Score:	6.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	10.71%	Indels:	0
DB:	2	Gaps:	0

US-09-980-043-4 (1-56) x US-08-925-171-3 (1-136)

```
        14 GluGlnValGlyHisAsp 19
           ||||||||||||||||||
        23 GAGCAGGTAGGCCATGAT 6

SULT 9
-09-243-335-3/c
Sequence 3, Application US/09243335A
Patent No. 6197580
GENERAL INFORMATION:
 APPLICANT: American Home Products Corp.
 APPLICANT: Susulic, Vedrana S.
 APPLICANT: Duzic, Edmir
 TITLE OF INVENTION: TRANSCRIPTIONAL REGULATION OF THE HUMAN
 TITLE OF INVENTION: B3-ADRENERGIC RECEPTOR GENE
 FILE REFERENCE: 0630/0E791
 CURRENT APPLICATION NUMBER: US/09/243,335A
 CURRENT FILING DATE: 1999-02-01
 NUMBER OF SEQ ID NOS: 49
 SOFTWARE: FastSEQ for Windows Version 3.0
SEQ ID NO 3
  LENGTH: 200
  TYPE: DNA
  ORGANISM: Homo sapien
  FEATURE:
-09-243-335-3

ignment Scores:
ed. No.:                109             Length:         200
ore:                    6.00            Matches:        6
rcent Similarity:       100.00%         Conservative:   0
st Local Similarity:    100.00%         Mismatches:     0
ery Match:              10.71%          Indels:         0
:                       3               Gaps:           0

-09-980-043-4 (1-56) x US-09-243-335-3 (1-200)

        45 LysGluGlyArgLysVal 50
           ||||||||||||||||||
       119 AAGGAAGGAAGGAAGGTC 102

SULT 10
-08-332-766A-14/c
Sequence 14, Application US/08332766A
Patent No. 5843647
 GENERAL INFORMATION:
   APPLICANT: JEFFREYS, Alec J.
   APPLICANT: ARMOUR, John
   TITLE OF INVENTION: SIMPLE TANDEM REPEATS
   NUMBER OF SEQUENCES: 125
   CORRESPONDENCE ADDRESS:
     ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.
     STREET: 1100 New York Avenue, N.W.
     CITY: Washington
     STATE: D. C.
     COUNTRY: U.S.A.
     ZIP: 20005-3918
   COMPUTER READABLE FORM:
     MEDIUM TYPE: Floppy disk
     COMPUTER: IBM PC compatible
     OPERATING SYSTEM: PC-DOS/MS-DOS
     SOFTWARE: PatentIn Release #1.0, Version #1.25
   CURRENT APPLICATION DATA:
     APPLICATION NUMBER: US/08/332,766A
     FILING DATE: 01-NOV-1994
     CLASSIFICATION: 435
   PRIOR APPLICATION DATA:
     APPLICATION NUMBER: GB 9326052.9
     FILING DATE: 21-DEC-1993
   ATTORNEY/AGENT INFORMATION:
     NAME: BIRD, Donald J.
     REGISTRATION NUMBER: 25,323
     REFERENCE/DOCKET NUMBER: 217211/M94/0434/GB
   TELECOMMUNICATION INFORMATION:
     TELEPHONE: (202) 861 5900
     TELEFAX: (202) 822-0944
     TELEX: 6714627 CUSH
 INFORMATION FOR SEQ ID NO: 14:
   SEQUENCE CHARACTERISTICS:
     LENGTH: 238 base pairs
     TYPE: nucleic acid
     STRANDEDNESS: single
     TOPOLOGY: linear
   MOLECULE TYPE: DNA (genomic)
US-08-332-766A-14

Alignment Scores:
Pred. No.:              129             Length:         238
Score:                  6.00            Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            10.71%          Indels:         0
DB:                     2               Gaps:           0

US-09-980-043-4 (1-56) x US-08-332-766A-14 (1-238)

Qy          44 IleLysGluGlyArgLys 49
               ||||||||||||| |||
Db         226 ATTAAGGAAGGAAGGAAG 209

RESULT 11
US-08-520-678A-25/c
; Sequence 25, Application US/08520678A
; Patent No. 5874565
; GENERAL INFORMATION:
;   APPLICANT: Rice, Charles M.
;   APPLICANT: Kolykhalov, Alexander A.
;   TITLE OF INVENTION: NOVEL 3' TERMINAL SEQUENCE OF HEPATITIS
;   TITLE OF INVENTION: C VIRUS GENOME AND DIAGNOSTIC AND THERAPEUTIC USES THEREOF
;   NUMBER OF SEQUENCES: 31
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Howell & Haferkamp, L.C.
;     STREET: 7733 Forsyth Blvd., Suite 1400
;     CITY: St. Louis
;     STATE: MO
;     COUNTRY: USA
;     ZIP: 63105
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/520,678A
;     FILING DATE:
;     CLASSIFICATION: 536
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Henderson, Melodie W.
;     REGISTRATION NUMBER: 37,848
;     REFERENCE/DOCKET NUMBER: 6029-6836
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: 314-727-5188
;     TELEFAX: 314 727-6092
;     TELEX:
; INFORMATION FOR SEQ ID NO: 25:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 253 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;   MOLECULE TYPE: RNA (genomic)
US-08-520-678A-25

Alignment Scores:
Pred. No.:              137             Length:         253
Score:                  6.00            Matches:        6
Percent Similarity:     100.00%         Conservative:   0
```

```
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            10.71%         Indels:        0
DB:                     2              Gaps:          0

US-09-980-043-4 (1-56) x US-08-520-678A-25 (1-253)

Qy          44 IleLysGluGlyArgLys 49
               |||||||||||||||||
Db         155 ATTAAAGAAGGAAGGAAA 138

RESULT 12
US-08-897-126-25/c
; Sequence 25, Application US/08897126
; Patent No. 6297003
;  GENERAL INFORMATION:
;    APPLICANT:  Rice, Charles M.
;    APPLICANT:  Kolykhalov, Alexander A.
;    TITLE OF INVENTION:  NOVEL 3' TERMINAL SEQUENCE OF HEPATITIS
;    TITLE OF INVENTION:  C VIRUS GENOME AND DIAGNOSTIC AND THERAPEUTIC USES THEREOF
;    NUMBER OF SEQUENCES:  39
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Howell & Haferkamp, L.C.
;      STREET:  7733 Forsyth Blvd., Suite 1400
;      CITY:  St. Louis
;      STATE:  MO
;      COUNTRY:  USA
;      ZIP:  63105
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/897,126
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/520,678
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Henderson, Melodie W.
;      REGISTRATION NUMBER:  37,848
;      REFERENCE/DOCKET NUMBER:  6029-6836
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  314-727-5188
;      TELEFAX:  314-727-6092
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  25:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  253 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  RNA (genomic)
US-08-897-126-25

Alignment Scores:
Pred. No.:              137            Length:        253
Score:                  6.00           Matches:       6
Percent Similarity:     100.00%        Conservative:  0
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            10.71%         Indels:        0
DB:                     3              Gaps:          0

US-09-980-043-4 (1-56) x US-08-897-126-25 (1-253)

Qy          44 IleLysGluGlyArgLys 49
               |||||||||||||||||
Db         155 ATTAAAGAAGGAAGGAAA 138

RESULT 13
US-08-520-678A-24/c
; Sequence 24, Application US/08520678A
; Patent No. 5874565
;  GENERAL INFORMATION:
;    APPLICANT:  Rice, Charles M.
;    APPLICANT:  Kolykhalov, Alexander A.
;    TITLE OF INVENTION:  NOVEL 3' TERMINAL SEQUENCE OF HEPATITIS
;    TITLE OF INVENTION:  C VIRUS GENOME AND DIAGNOSTIC AND THERAPEUTIC USES THEREOF
;    NUMBER OF SEQUENCES:  39
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Howell & Haferkamp, L.C.
;      STREET:  7733 Forsyth Blvd., Suite 1400
;      CITY:  St. Louis
;      STATE:  MO
;      COUNTRY:  USA
;      ZIP:  63105
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/520,678A
;      FILING DATE:
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Henderson, Melodie W.
;      REGISTRATION NUMBER:  37,848
;      REFERENCE/DOCKET NUMBER:  6029-6836
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  314-727-5188
;      TELEFAX:  314-727-6092
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  24:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  257 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-520-678A-24

Alignment Scores:
Pred. No.:              139            Length:        257
Score:                  6.00           Matches:       6
Percent Similarity:     100.00%        Conservative:  0
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            10.71%         Indels:        0
DB:                     2              Gaps:          0

US-09-980-043-4 (1-56) x US-08-520-678A-24 (1-257)

Qy          44 IleLysGluGlyArgLys 49
               |||||||||||||||||
Db         221 ATTAAAGAAGGAAGAAAA 204

RESULT 14
US-08-897-126-24/c
; Sequence 24, Application US/08897126
; Patent No. 6297003
;  GENERAL INFORMATION:
;    APPLICANT:  Rice, Charles M.
;    APPLICANT:  Kolykhalov, Alexander A.
;    TITLE OF INVENTION:  NOVEL 3' TERMINAL SEQUENCE OF HEPATITIS
;    TITLE OF INVENTION:  C VIRUS GENOME AND DIAGNOSTIC AND THERAPEUTIC USES THEREOF
;    NUMBER OF SEQUENCES:  39
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Howell & Haferkamp, L.C.
;      STREET:  7733 Forsyth Blvd., Suite 1400
;      CITY:  St. Louis
;      STATE:  MO
;      COUNTRY:  USA
;      ZIP:  63105
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/897,126
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/520,678
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Henderson, Melodie W.
REGISTRATION NUMBER: 37,848
REFERENCE/DOCKET NUMBER: 6029-6836
TELECOMMUNICATION INFORMATION:
TELEPHONE: 314-727-5188
TELEFAX: 314-727-6092
TELEX:
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 257 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
-08-897-126-24

ignment Scores:

ed. No.:	139	Length:	257
ore:	6.00	Matches:	6
rcent Similarity:	100.00%	Conservative:	0
st Local Similarity:	100.00%	Mismatches:	0
ery Match:	10.71%	Indels:	0
:	3	Gaps:	0

-09-980-043-4 (1-56) x US-08-897-126-24 (1-257)

```
         44 IleLysGluGlyArgLys 49
            ||||||||||||||||||
        221 ATTAAAGAAGGAAGAAAA 204
```

SULT 15
-09-313-294A-2372
Sequence 2372, Application US/09313294A
Patent No. 6476212
GENERAL INFORMATION:
APPLICANT: Lalgudi, Raghunath V.
APPLICANT: Ito, Laura Y.
APPLICANT: Sherman, Bradley K.
TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
FILE REFERENCE: PL-0017 US
CURRENT APPLICATION NUMBER: US/09/313,294A
CURRENT FILING DATE: 1999-05-14
NUMBER OF SEQ ID NOS: 7600
SOFTWARE: PERL Program
SEQ ID NO 2372
LENGTH: 261
TYPE: DNA
ORGANISM: Zea mays
FEATURE:
NAME/KEY: misc_feature
OTHER INFORMATION: Incyte ID No. 6476212 700552566H1
-09-313-294A-2372

ignment Scores:

ed. No.:	141	Length:	261
ore:	6.00	Matches:	6
rcent Similarity:	100.00%	Conservative:	0
st Local Similarity:	100.00%	Mismatches:	0
ery Match:	10.71%	Indels:	0
.	4	Gaps:	0

-09-980-043-4 (1-56) x US-09-313-294A-2372 (1-261)

```
Qy          37 LysSerCysPheGluPhe 42
               |||||||||||||||||||
Db         118 AAATCATGTTTTGAATTT 135
```

Search completed: November 5, 2003, 04:48:35
Job time : 10.9362 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 31, 2003, 17:45:37 ; Search time 41.2147 Seconds
(without alignments)
3667.830 Million cell updates/sec

Title: US-09-980-043-4
Perfect score: 56
Sequence: 1 MMESGAPICHTCGEQVGHDA.....KSCFEFEIKEGRKVCLRCGS 56

Scoring table: OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Word size: 1

Total number of hits satisfying chosen parameters: 5103488

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09980043/runat_29102003_163930_22655/app_query.fasta_1.1358
-DB=N_Geneseq_19Jun03 -QFMT=fastap -SUFFIX=oligo.rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09980043_@CGN_1_1_458_@runat_29102003_163930_22655 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : N_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		56	100.0	1010	22	AAC65449	Populus tremuloide
2		44	78.6	3232	22	AAC65448	Populus tremuloide
3		8	14.3	736	21	AAA67115	Pinus radiata cell
4		7	12.5	133	25	ABX28011	Human GDP-mannose
5	c	7	12.5	198	22	AAH52138	Human AFP protein
6		7	12.5	235	19	AAX14004	H. pylori GHPO 131
7	c	7	12.5	319	24	ABN64906	Human cancer relat
8		7	12.5	467	24	ABK80286	Bacillus clausii g
9	c	7	12.5	475	23	ABV45889	Human prostate exp
10		7	12.5	507	21	AAC56867	Pinus radiata tran
11	c	7	12.5	552	23	ABV45963	Human prostate exp
12	c	7	12.5	561	22	AAK61767	Human immune/haema
13		7	12.5	651	24	ABK33134	DNA encoding novel
14		7	12.5	664	21	AAA67127	Eucalyptus grandis
15	c	7	12.5	670	22	AAK52619	Human polynucleoti
16		7	12.5	848	24	ABK75624	Bacillus lichenifo
17		7	12.5	856	21	AAA67079	Eucalyptus grandis
18		7	12.5	885	21	AAZ29261	Rifin 3208 gene.
19		7	12.5	939	21	AAA70087	Plasmodium falcipa
20		7	12.5	1024	24	ABX66751	Helicobacter pylor
21	c	7	12.5	1704	22	AAK51635	Human polynucleoti
22	c	7	12.5	1914	21	AAA96963	Nucleotide sequenc
23		7	12.5	2187	23	ABL17227	Drosophila melanog
24		7	12.5	2459	23	ABL11985	Drosophila melanog
25	c	7	12.5	3164	23	ABL10437	Drosophila melanog
26		7	12.5	3207	19	AAV08372	Cellulose synthase
27		7	12.5	3328	19	AAV34432	Cotton cellulose s
28		7	12.5	3747	21	AAA67145	Eucalyptus grandis
29		7	12.5	3851	21	AAA67114	Pinus radiata cell
30		7	12.5	5541	23	ABL17226	Drosophila melanog
31		7	12.5	6510	22	AAS46647	Tumour suppressor
32		7	12.5	6510	24	ABL32394	Human immune syste
33		7	12.5	6510	24	AAS61114	Human gene regulat
34	c	7	12.5	8259	23	ABL16826	Drosophila melanog
35	c	7	12.5	8372	23	ABL10436	Drosophila melanog
36		7	12.5	8688	21	AAC42748	Arabidopsis thalia
37		7	12.5	8688	24	ABZ13943	Arabidopsis thalia
38	c	7	12.5	9355	23	ABL11984	Drosophila melanog
39		7	12.5	14924	24	ABL54322	Chemically treated
40		7	12.5	14924	24	ABL32225	Human immune syste
41	c	7	12.5	66109	24	ABN97273	Gene #3771 used to
42	c	7	12.5	66109	24	ABL62723	Colon adenocarcino
43	c	7	12.5	109906	24	ABK94411	DNA encoding endot
44	c	7	12.5	122748	24	ABT10719	Human breast cance
45	c	7	12.5	198285	24	ABK84699	Human cDNA differe

ALIGNMENTS

RESULT 1
AAC65449
ID AAC65449 standard; DNA; 1010 BP.
XX
AC AAC65449;
XX
DT 07-MAR-2001 (first entry)
XX
DE Populus tremuloides CelA 5' flanking region.
XX
KW Aspen; CelA; cellulose synthase; arabidopsis; wood;
KW mechanical stress response element; MSRE; cellulose biosynthesis; ds.
XX
OS Populus tremuloides.
XX
PN WO200071670-A2.
XX

30-NOV-2000.

18-MAY-2000; 2000WO-US13637.

21-MAY-1999; 99US-0135280.

(UNMT) UNIV MICHIGAN TECHNOLOGICAL.

Chiang VL, Wu L, Chandrashekhar JP, Carraway DT;

WPI; 2001-032025/04.
P-PSDB; AAB37894.

Novel polynucleotide encoding cellulose synthase useful for enhancing cellulose biosynthesis, modifying lignin biosynthesis, and for accelerating growth of plants and for upregulating cellulose synthase levels -

Claim 2; Fig 4; 65pp; English.

The present sequence is given in a specification relating to polynucleotides encoding cellulose synthase enzymes from Populus tremuloides and Arabidopsis. The cellulose synthases are designated PtCelA and AraxCelA respectively. The polynucleotides may be introduced into plants where they are useful for altering the growth of a plant. Vectors comprising a cellulose synthase promoter, or a fragment of the promoter containing one or more mechanical stress response elements (MSRE), operably linked to a cellulose synthase gene are useful for causing stress-induced gene expression in a plant cell. A cDNA encoding a protein that binds to a positive MSRE may be delivered into the plant to increase expression of cellulose synthase. A cDNA in an antisense orientation may be delivered to reduce expression of cellulose synthase. The polynucleotides enable the development of transgenic tree varieties having increased cellulose content, decreased lignin content and therefore improved wood fibre characteristics. The cellulose synthase promoters are useful for synthesising greater quantities of high-crystalline cellulose in plant and preferably in trees, which permits an increase in proportion of cellulose in transgenic plants, greater strength of juvenile wood and fiber and acceleration of overall growth rate.

Sequence 1010 BP; 289 A; 212 C; 216 G; 293 T; 0 other;

```
ignment Scores:
ed. No.:                1.55e-53        Length:         1010
ore:                    56.00           Matches:        56
rcent Similarity:       100.00%         Conservative:   0
st Local Similarity:    100.00%         Mismatches:     0
ery Match:              100.00%         Indels:         0
:                       22              Gaps:           0

-09-980-043-4 (1-56) x AAC65449 (1-1010)

          1 MetMetGluSerGlyAlaProIleCysHisThrCysGlyGluGlnValGlyHisAspAla 20
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
        841 ATGATGGAATCTGGGGCTCCTATATGCCATACCTGTGGTGAACAGGTGGGGCATGATGCA 900

         21 AsnGlyGluLeuPheValAlaCysHisGluCysSerTyrProMetCysLysSerCysPhe 40
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
        901 AATGGGGAGCTATTTGTGGCTTGCCATGAGTGTAGCTATCCCATGTGCAAGTCTTGTTTC 960

         41 GluPheGluIleLysGluGlyArgLysValCysLeuArgCysGlySer 56
            ||||||||||||||||||||||||||||||||||||||||||||||||||
        961 GAGTTTGAAATCAAAGAGGGCCGGAAAGTTTGCTTGCGGTGTGGCTCG 1008

SULT 2
C65448
   AAC65448 standard; cDNA; 3232 BP.

   AAC65448;

   07-MAR-2001 (first entry)

DE   Populus tremuloides Cel? cDNA.
XX
KW   Aspen; CelA; cellulose synthase; Arabidopsis; wood;
KW   mechanical stress response element; MSRE; cellulose biosynthesis; ss
XX
OS   Populus tremuloides.
XX
PN   WO200071670-A2.
XX
PD   30-NOV-2000.
XX
PF   18-MAY-2000; 2000WO-US13637.
XX
PR   21-MAY-1999;  99US-0135280.
XX
PA   (UNMT ) UNIV MICHIGAN TECHNOLOGICAL.
XX
PI   Chiang VL, Wu L, Chandrashekhar JP, Carraway DT;
XX
DR   WPI; 2001-032025/04.
DR   P-PSDB; AAB37893.
XX
PT   Novel polynucleotide encoding cellulose synthase useful for enhancing
PT   cellulose biosynthesis, modifying lignin biosynthesis, and for
PT   accelerating growth of plants and for upregulating cellulose synthase
PT   levels -
XX
PS   Claim 1; Fig 1; 65pp; English.
XX
CC   The present sequence is given in a specification relating to
CC   polynucleotides encoding cellulose synthase enzymes from Populus
CC   tremuloides and Arabidopsis. The cellulose synthases are designated
CC   PtCelA and AraxCelA respectively. The polynucleotides may be introduced
CC   into plants where they are useful for altering the growth of a plant.
CC   Vectors comprising a cellulose synthase promoter, or a fragment of the
CC   promoter containing one or more mechanical stress response elements
CC   (MSRE), operably linked to a cellulose synthase gene are useful for
CC   causing stress-induced gene expression in a plant cell. A cDNA encoding a
CC   protein that binds to a positive MSRE may be delivered into the plant to
CC   increase expression of cellulose synthase. A cDNA in an antisense
CC   orientation may be delivered to reduce expression of cellulose synthase.
CC   The polynucleotides enable the development of transgenic tree varieties
CC   having increased cellulose content, decreased lignin content and
CC   therefore improved wood fibre characteristics. The cellulose synthase
CC   promoters are useful for synthesising greater quantities of
CC   high-crystalline cellulose in plant and preferably in trees, which
CC   permits an increase in proportion of cellulose in transgenic plants,
CC   greater strength of juvenile wood and fiber and acceleration of overall
CC   growth rate.
XX
SQ   Sequence 3232 BP; 895 A; 668 C; 772 G; 897 T; 0 other;

Alignment Scores:
Pred. No.:              1.33e-39        Length:         3232
Score:                  44.00           Matches:        44
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            78.57%          Indels:         0
DB:                     22              Gaps:           0

US-09-980-043-4 (1-56) x AAC65448 (1-3232)

Qy          1 MetMetGluSerGlyAlaProIleCysHisThrCysGlyGluGlnValGlyHisAspAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         69 ATGATGGAATCTGGGGCTCCTATATGCCATACCTGTGGTGAACAGGTGGGGCATGATGCA 128

Qy         21 AsnGlyGluLeuPheValAlaCysHisGluCysSerTyrProMetCysLysSerCysPhe 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        129 AATGGGGAGCTATTTGTGGCTTGCCATGAGTGTAGCTATCCCATGTGCAAGTCTTGTTTC 188

Qy         41 GluPheGluIle 44
              ||||||||||||
Db        189 GAGTTTGAAATC 200
```

```
RESULT 3
AAA67115
ID   AAA67115 standard; DNA; 736 BP.
XX
AC   AAA67115;
XX
DT   31-OCT-2000  (first entry)
XX
DE   Pinus radiata cellulose synthase nucleotide sequence SEQ ID NO:70.
XX
KW   Eucalyptus grandis; pinus radiata; Monterey pine; modification;
KW   plant cell wall; polysaccharide; polysaccharide biosynthetic pathway;
KW   transgenic plant; ds.
XX
OS   Pinus radiata.
XX
PN   WO200022092-A2.
XX
PD   20-APR-2000.
XX
PF   08-OCT-1999;   99WO-NZ00169.
XX
PR   13-OCT-1998;   98US-0170862.
PR   11-AUG-1999;   99US-0148426.
XX
PA   (GENE-) GENESIS RES & DEV CORP LTD.
PA   (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI   Bloksberg LN;
XX
DR   WPI; 2000-339328/29.
DR   P-PSDB; AAB16308.
XX
PT   New genes encoding proteins involved in a plant polysaccharide
PT   biosynthetic pathway, useful for modulating or altering the
PT   polysaccharide content, composition or structure of the plant -
XX
PS   Claim 1; Page 72; 301pp; English.
XX
CC   The present invention describes isolated polynucleotides (PN) comprising
CC   a sequence selected from one of 835 nucleotide sequences given in
CC   AAA67073 to AAA67907, their (reverse) complements, sequences producing
CC   an Expectation (E) value of 0.01 or less compared to the 835 sequences,
CC   sequences at least 50% identical to them, 200, 100, 40 or 20-mers of the
CC   835 sequences or sequences that are degenerately equivalent or allelic
CC   to the 835 sequences. The polynucleotides are used to modify the
CC   activity of a polypeptide involved in a polysaccharide biosynthetic
CC   pathway in the plant. They are especially used to modulate or alter the
CC   polysaccharide content, composition or structure of the plant. AAB16268
CC   to AAB16340 are proteins encoded by part of the polynucleotide sequence
CC   given in the present invention.
XX
SQ   Sequence 736 BP; 216 A; 163 C; 210 G; 147 T, 0 other;

Alignment Scores:
Pred. No.:              8.95          Length:          736
Score:                  8.00          Matches:           8
Percent Similarity:   100.00%         Conservative.      0
Best Local Similarity: 100.00%        Mismatches:        0
Query Match:           14.29%         Indels:            0
DB:                    21             Gaps:              0

US-09-980-043-4 (1-56) x AAA67115 (1-736)

Qy         21 AsnGlyGluLeuPheValAlaCys 28
              ||||||||||||||||||||||||
Db        193 AATGGAGAACTATTCGTGGCCTGC 216

RESULT 4
ABX28011
ID   ABX28011 standard; cDNA; 133 BP.
XX
AC   ABX28011;
XX
DT   11-FEB-2003  (first entry)
XX
DE   Human GDP-mannose 4,6-dehydratase (GM4,6D) DNA #10068.
XX
KW   Human; GDP-mannose 4,6-dehydratase; GM4,6D; gene; ss; inflammation;
KW   cellular fucosylation; glycoconjugate fucosylation; transplant rejection;
KW   arthritis; asthma; sepsis; reperfusion injury; stroke; infection;
KW   complex carbohydrate; gene replacement therapy; immunosuppressive;
KW   antiinflammatory; antiarthritic; antibacterial; cerebroprotective;
KW   antiasthmatic; vasotropic.
XX
OS   Homo sapiens.
XX
PN   US2002110548-A1.
XX
PD   15-AUG-2002.
XX
PF   11-JUN-2001; 2001US-0878574.
XX
PR   22-NOV-1996;   96US-0753233.
PR   03-DEC-1997;   97US-0984246.
PR   09-SEP-1998;   98US-0149674.
PR   14-JUN-1999;   99US-0333177.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Sullivan F,  Kriz R,  Kumar R;
XX
DR   WPI; 2003-066673/06.
XX
PT   New composition comprising GDP-mannose 4,6-dehydratase (GM4,6D)
PT   peptide, for manufacturing complex carbohydrates, or as targets for
PT   screening GM4,6D antagonists for treating e.g. arthritis, or transplant
PT   rejection -
XX
PS   Disclosure; SEQ ID NO 10070; 6pp; English.
XX
CC   The invention relates to a composition comprising a human GDP-mannose
CC   4,6-dehydratase (GM4,6D) peptide. The peptide is useful for identifying
CC   GM4,6D inhibitors. GM4,6D inhibitors are useful for reducing inflammation
CC   in a mammalian subject and for treating or ameliorating diseases affected
CC   by the level of cellular fucosylation or diseases affected by the
CC   fucosylation of glycoconjugates. These diseases include arthritis,
CC   transplant rejection, asthma, sepsis, reperfusion injury, stroke or
CC   infection. The GM4,6D peptide or a polynucleotide encoding it is also
CC   useful for manufacturing complex carbohydrates and as targets for
CC   screening small molecule antagonists of the activity of the enzyme. The
CC   polynucleotide is useful in developing an assay for defects in the
CC   enzyme, as well as in gene replacement therapy. Sequences
CC   ABX17942-ABX17944 and ABX17947-ABX33716 represent DNA molecules encoding
CC   human GM4,6D peptides of the invention.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification but was obtained in electronic format directly from USPTO
CC   at seqdata.uspto.gov/sequence.html.
XX
SQ   Sequence 133 BP; 36 A; 31 C; 38 G; 28 T; 0 other;

Alignment Scores:
Pred. No.:              25            Length:          133
Score:                  7.00          Matches:           7
Percent Similarity:   100.00%         Conservative:      0
Best Local Similarity: 100.00%        Mismatches:        0
Query Match:           12.50%         Indels:            0
DB:                    25             Gaps:              0

US-09-980-043-4 (1-56) x ABX28011 (1-133)

Qy         43 GluIleLysGluGlyArgLys 49
              |||||||||||||||||||||
Db         36 GAGATCAAGGAAGGAAGGAAG 56
```

```
:SULT 5
AH52138/c
)   AAH52138 standard; cDNA; 198 BP.
(
:   AAH52138;
(
`   10-SEP-2001  (first entry)
(
:   Human AFP protein encoding cDNA sequence SEQ ID NO:91.
(
¥   Human; secreted protein; secretion; bacterial cell; fungal cell;
¥   eukaryotic cell; fusion protein; maltose binding protein;
¥   immunoglobulin constant region; polyhistidine tag; ss.
(
3   Homo sapiens.
(
¥   WO200129221-A2.
(
)   26-APR-2001.
(
=   20-OCT-2000; 2000WO-US29052.
(
?   20-OCT-1999;  99US-0160712.
(
\   (ZYMO ) ZYMOGENETICS INC.
(
    Conklin DC,  Yee DP;
(
?   WPI; 2001-300340/31.
?   P-PSDB; AAG81287.
(
    Isolated polypeptide for directing secretion of proteins of interest
    from a host cell including, e.g. bacteria, includes contiguous amino
    acid residues of polypeptide with specified amino acids -
(
>   Claim 9; Page 201; 617pp; English.

`   AAH52093 to AAH52303 encode the human secreted proteins given in AAG81242
:   to AAG81453. The secreted proteins can be used for directing the
`   secretion of proteins of interest from a host cell including bacteria,
`   fungal cells, and cultured higher eukaryotic cells. The present invention
`   also describes fusion proteins, where a secreted protein of the invention
`   is operably linked via a peptide bond or peptide linker to a second
`   protein selected from the group consisting of maltose binding protein,
`   an immunoglobulin constant region, a polyhistidine tag and a peptide
`   given in AAG81453.

)   Sequence 198 BP; 28 A; 70 C; 62 G; 37 T; 1 other;

ignment Scores:
ed. No.:                35.9          Length:         198
ore:                    7.00          Matches:        7
rcent Similarity:       100.00%       Conservative:   0
st Local Similarity:    100.00%       Mismatches:     0
ery Match:              12.50%        Indels:         0
3:                      22            Gaps:           0

 -09-980-043-4 (1-56) x AAH52138 (1-198)

/         14 GluGlnValGlyHisAspAla 20
             ||||||||||||||||||||
>        165 GAGCAGGTCGGGCACGACGCC 145

:SULT 6
AX14004
)   AAX14004 standard; DNA; 235 BP.

`   AAX14004;

`   31-MAR-1999  (first entry)

    H. pylori GHPO 1316 gene.

KW   GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;
KW   peptic ulcer disease; ss.
XX
OS   Helicobacter pylori.
XX
FH   Key             Location/Qualifiers
FT   CDS             51..185
FT                   /*tag=
XX
PN   WO9843478-A1.
XX
PD   08-OCT-1998.
XX
PF   01-APR-1998;    98WO-US06471
XX
PR   29-JUL-1997;    97US-0906415
PR   01-APR-1997;    97US-0841357
PR   24-JUN-1997;    97US-0881127
XX
PA   (HUMA-) HUMAN GENOME S[illegible] INC.
PA   (INMR ) MERIEUX CRAVAX (PASTEUR MERIEUX SERUMS.
XX
PI   Al-Garawi A,  Kleanthous H,  Miller C,  Oomen RP,  Tomb J;
XX
DR   WPI; 1998-542293/46.
DR   P-PSDB; AAW98285.
XX
PT   New isolated Helicobacter polynucleotides - used to develop products
PT   for the diagnosis, prevention and treatment of Helicobacter
PT   infections and gastrointestinal diseases
XX
PS   Claim 1; Page 338; 2054pp; English.
XX
CC   This sequence represents a polynucleotide of the invention. It was
CC   isolated from Helicobacter pylori and encodes a H.pylori GHPO protein.
CC   The polypeptides can be used for preventing or treating Helicobacter
CC   infections, and gastroduodenal diseases associated with these
CC   infections, including acute, chronic, and atrophic gastritis, and peptic
CC   ulcer diseases, e.g. gastric and duodenal ulcers. They can also be used
CC   for the production of antibodies. The products can also be used for
CC   detection and diagnosis
XX
SQ   Sequence 235 BP; 94 A; 48 C; 31 G; 72 T; 0 other;

Alignment Scores:
Pred. No.:                41.9          Length:         235
Score:                    7.00          Matches:        7
Percent Similarity:       100.00%       Conservative:   0
Best Local Similarity:    100.00%       Mismatches:     0
Query Match:              12.50%        Indels:         0
DB:                       19            Gaps:           0

US-09-980-043-4 (1-56) x AAX14004 (1-235)

Qy          24 LeuPheValAlaCysHisGlu 30
               ||||||||||||||||||||||
Db          67 TTGTTTCTGGCTTGCACGAA 107

RESULT 7
ABN64906/c
ID   ABN64906 standard; cDNA  319 BP.
XX
AC   ABN64906;
XX
DT   28-JUN-2002  (first entry)
XX
DE   Human cancer related polynucleotide SEQ ID NO 4873.
XX
KW   Human; cytostatic; gene expression; gene mapping; tissue profiling;
KW   gene therapy; cancer; tumour; gene; ss.
XX
OS   Homo sapiens.
XX
```

PN WO200214500-A2.
XX
PD 21-FEB-2002.
XX
PF 16-AUG-2001; 2001WO-US25840.
XX
PR 16-AUG-2000; 2000US-226326P.
XX
PA (CHIR) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX
PI Escobedo J, Garcia PD, Suddath Klinger J, Reinhard C, Randazzo F;
PI Lamson G, Scott EM, Zhang G, Kassam A, Pot D, Labat I;
XX
DR WPI; 2002-241905/29.
XX
PT New nucleic acid for producing a polypeptide, detecting differentially
PT expressed genes correlated with a cancerous state of a mammalian cell,
PT and inhibiting tumor growth -
XX
PS Claim 1; SEQ ID NO 4873; 883pp + Sequence Listing; English.
XX
CC The invention relates to an isolated polynucleotide (ABN27253-ABN33262)
CC with cytostatic activity. The polynucleotide is used to produce a
CC polypeptide, to detect differentially expressed genes correlated with a
CC cancerous state of a mammalian cell and to inhibit tumour growth. The
CC polynucleotide is used as a probe in mapping and tissue profiling. The
CC encoded polypeptide and antibodies to the polypeptide can also be used
CC for therapeutic and diagnostic purposes. The polynucleotide is useful for
CC gene therapy.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 319 BP; 89 A; 78 C; 50 G; 101 T; 1 other;

Alignment Scores:

Pred. No.:	55.4	Length:	319
Score:	7.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	12.50%	Indels:	0
DB:	24	Gaps:	0

US-09-980-043-4 (1-56) x ABN64906 (1-319)

```
Qy         43 GluIleLysGluGlyArgLys 49
              |||||||||||||||||||
Db        152 GAAATAAAGGAGGGGAGAAAA 172
```

RESULT 8
ABK80286
ID ABK80286 standard; DNA; 467 BP.
XX
AC ABK80286;
XX
DT 13-AUG-2002 (first entry)
XX
DE Bacillus clausii genomic sequence tag (GST) #3129.
XX
KW Differential gene expression; genomic sequenced tag; GST;
KW altered culture condition; environmental stress;
KW physiological provocation; ds.
XX
OS Bacillus clausii.
XX
PN WO200229113-A2.
XX
PD 11-APR-2002.
XX
PF 05-OCT-2001; 2001WO-US31437.
XX
PR 06-OCT-2000; 2000US-0680598.
PR 27-MAR-2001; 2001US-279526P.
XX
PA (NOVO) NOVOZYMES BIOTECH INC.
PA (NOVO) NOVOZYMES AS.
XX
PI Berka R, Clausen IG;
XX
DR WPI; 2002-416684/44.
XX
PT Monitoring differential expression of several genes in first Bacillus
PT cell relative to expression of same genes in one or more second
PT Bacillus cells, by using substrate containing Bacillus genomic
PT sequenced tag array -
XX
PS Claim 11; SEQ ID NO 7577; 200pp; English.
XX
CC The invention describes a method of monitoring differential expression of
CC genes in a first Bacillus cell relative to expression of the genes in
CC other Bacillus cells, comprising hybridising labelled nucleic acid probes
CC isolated from Bacillus cells to a substrate containing array of Bacillus
CC genomic sequenced tags (GST), examining the array, and determining
CC relative gene expression by an observed hybridisation reporter signal of
CC a spot in the array. The method is useful for measuring the expression of
CC genes in a first Bacillus cell relative to expression of the same genes
CC in one or more second Bacillus cells. The method is useful for monitoring
CC global expression of several genes from a Bacillus cell, discovering new
CC genes, identifying possible functions of unknown open reading frames and
CC monitoring gene copy number variation and stability. Monitoring changes
CC in expression of genes may be used to provide a representation of the way
CC in which Bacillus cells adapt to changes in culture conditions,
CC environmental stress or other physiological provocation. Extensive
CC follow-up characterisation is unnecessary, when one spot on an array
CC equals one gene or one open reading frame, since sequence information is
CC available. This sequence represents a genomic sequence tag (GST) used in
CC the method of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 467 BP; 126 A; 92 C; 129 G; 120 T; 0 other;

Alignment Scores:

Pred. No.:	78.4	Length:	467
Score:	7.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	12.50%	Indels:	0
DB:	24	Gaps:	0

US-09-980-043-4 (1-56) x ABK80286 (1-467)

```
Qy         44 IleLysGluGlyArgLysVal 50
              |||||||||||||||||||||
Db        397 ATAAAGGAAGGCCGGAAAGTC 417
```

RESULT 9
ABV45889/c
ID ABV45889 standard; cDNA; 475 BP.
XX
AC ABV45889;
XX
DT 16-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 45880.
XX
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodyanamic marker;
KW pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200160860-A2.
XX

23-AUG-2001.

20-FEB-2001; 2001WO-US05171.

17-FEB-2000; 2000US-183319P.
16-MAR-2000; 2000US-189862P.
25-MAY-2000; 2000US-207454P.
09-JUN-2000; 2000US-211314P.
18-JUL-2000; 2000US-219007P.
13-DEC-2000; 2000US-255281P.

(MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

Schlegel R, Endege WO, Monahan JE;

WPI; 2001-662795/76.

Novel isolated nucleic acid molecule associated with cancerous state of prostate cells and correlating with presence of prostate cancer, useful for detecting presence of prostate cancer, stage of prostate cancer

Claim 1; Page 9069; 11750pp; English.

The invention relates to an isolated nucleic acid molecule (I) comprising a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the specification or its complement. (I) is useful for:
(a) assessing whether a patient is afflicted with prostate cancer;
(b) monitoring the progression of prostate cancer in a patient;
(c) assessing the efficacy of a test compound to inhibit prostate cancer in a patient;
(d) assessing the efficacy of a therapy for inhibiting prostate cancer in a patient;
(e) selecting a composition for inhibiting prostate cancer in a patient;
(f) assessing the prostate cell carcinogenic potential of a compound;
(g) determining whether prostate cancer has metastasized in a patient;
(h) assessing the aggressiveness or indolence of prostate cancer in a patient;
(I) is also useful as a pharmacodyanamic or pharmacogenomic marker.

Sequence 475 BP; 106 A; 121 C; 147 G; 101 T; 0 other;

ignment Scores:
ed. No.: 79.7 Length: 475
ore: 7.00 Matches: 7
rcent Similarity: 100.00% Conservative: 0
st Local Similarity: 100.00% Mismatches: 0
ery Match: 12.50% Indels: 0
: 23 Gaps: 0

-09-980-043-4 (1-56) x ABV45889 (1-475)

5 GlyAlaProIleCysHisThr 11
||||||||||||||||||||
31 GGAGCTCCTATTTGCCACACA 11

SULT 10
C56867
AAC56867 standard; DNA; 507 BP.

AAC56867;

25-JAN-2001 (first entry)

Pinus radiata transcription factor DNA sequence #313.

Plant; transcription factor; gene expression; eucalyptus; pine; acacia; poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor; basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS; homeodomain zipper; LIM domain; AP2; EREBs; zinc finger domain; type 2 Cys2His2; CCAAT box element; MYB; ss.

Pinus radiata.

PN WO200053724-A2.
XX
PD 14-SEP-2000.
XX
PF 09-MAR-2000; 2000WO-US06112.
XX
PR 11-MAR-1999; 99US-026[illegible]513.
PR 18-AUG-1999; 99US-014[illegible]485.
XX
PA (GENE-) GENESIS RES & DEV CORP LTD.
PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI Wood M, McGrath A, Shenk MA, Glenn M;
XX
DR WPI; 2000-579369/54.
XX
PT New isolated polynucleotide encoding a plant transcription factor for
PT producing a plant e.g. woody plant, preferably eucalyptus or pine,
PT having modified gene expression or modified activity of a polypepride
PT -
XX
PS Claim 1; Page 533; 747pp; English.
XX
CC The present invention relates to novel plant transcription factors from
CC Eucalyptus grandis or Pinus radiata. The present sequence is the coding
CC sequence for one such transcription factor. The transcription factor may
CC be used to produce a plant having modified gene expression such as a
CC woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or
CC mahogany species or to modify the activity of a polypeptide in a plant.
CC The transcription factors of the present invention are members from the
CC following families of regulatory proteins: bZIP, bZIP family of G-box
CC binding factors, basic helix-loop-helix zipper,
CC homeotic/homeodomain/homeobox/MADS, homeodomain zipper, LIM domain, AP2
CC and EREBs, zinc finger domains of type 2 Cys2His2, CCAAT box elements
CC and MYB.
XX
SQ Sequence 507 BP; 128 A; 118 C; 135 G; 126 T; 0 other;

Alignment Scores:
Pred. No.: 84.5 Length: 507
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 12.50% Indels: 0
DB: 21 Gaps: 0

US-09-980-043-4 (1-56) x AAC56867 (1-507)

Qy 7 ProIleCysHisThrCysGly 13
||||||||||||||||||||
Db 23 CCCATTTGTCACACATGCGGC 43

RESULT 11
ABV45963/c
ID ABV45963 standard; cDNA; 552 BP.
XX
AC ABV45963;
XX
DT 16-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 45954.
XX
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodyanamic marker;
KW pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200160860-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US05171.
XX

PR 17-FEB-2000; 2000US-183319P.
PR 16-MAR-2000; 2000US-189862P.
PR 25-MAY-2000; 2000US-207454P.
PR 09-JUN-2000; 2000US-211314P.
PR 18-JUL-2000; 2000US-219007P.
PR 13-DEC-2000; 2000US-255281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JE;
XX
DR WPI; 2001-662795/76.
XX
PT Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer -
XX
PS Claim 1; Page 9082; 11750pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX
SQ Sequence 552 BP; 128 A; 146 C; 133 G; 145 T; 0 other;

Alignment Scores:

Pred. No.:	91.4	Length:	552
Score:	7.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	12.50%	Indels:	0
DB:	23	Gaps:	0

US-09-980-043-4 (1-56) x ABV45963 (1-552)

Qy 5 GlyAlaProIleCysHisThr 11
Db 32 GGAGCTCCTATTTGCCACA 50

RESULT 12
AAK61767/c
ID AAK61767 standard; cDNA; 561 BP.
XX
AC AAK61767;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen encoding cDNA SEQ ID NO:6827.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ss.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09 AUG-2001.
XX
PF 17 JAN-2001; 2001WO-US01354
XX
PR 31-JAN-2000; 2000US-0179065
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.

02-OCT-2000; 2000US-0237038.
02-OCT-2000; 2000US-0237039.
02-OCT-2000; 2000US-0237040.
13-OCT-2000; 2000US-0239935.
13-OCT-2000; 2000US-0239937.
20-OCT-2000; 2000US-0240960.
20-OCT-2000; 2000US-0241221.
20-OCT-2000; 2000US-0241785.
20-OCT-2000; 2000US-0241786.
20-OCT-2000; 2000US-0241787.
20-OCT-2000; 2000US-0241808.
20-OCT-2000; 2000US-0241809.
20-OCT-2000; 2000US-0241826.
01-NOV-2000; 2000US-0244617.
08-NOV-2000; 2000US-0246474.
08-NOV-2000; 2000US-0246475.
08-NOV-2000; 2000US-0246476.
08-NOV-2000; 2000US-0246477.
08-NOV-2000; 2000US-0246478.
08-NOV-2000; 2000US-0246523.
08-NOV-2000; 2000US-0246524.
08-NOV-2000; 2000US-0246525.
08-NOV-2000; 2000US-0246526.
08-NOV-2000; 2000US-0246527.
08-NOV-2000; 2000US-0246528.
08-NOV-2000; 2000US-0246532.
08-NOV-2000; 2000US-0246609.
08-NOV-2000; 2000US-0246610.
08-NOV-2000; 2000US-0246611.
08-NOV-2000; 2000US-0246613.
17-NOV-2000; 2000US-0249207.
17-NOV-2000; 2000US-0249208.
17-NOV-2000; 2000US-0249209.
17-NOV-2000; 2000US-0249210.
17-NOV-2000; 2000US-0249211.
17-NOV-2000; 2000US-0249212.
17-NOV-2000; 2000US-0249213.
17-NOV-2000; 2000US-0249214.
17-NOV-2000; 2000US-0249215.
17-NOV-2000; 2000US-0249216.
17-NOV-2000; 2000US-0249217.
17-NOV-2000; 2000US-0249218.
17-NOV-2000; 2000US-0249244.
17-NOV-2000; 2000US-0249245.
17-NOV-2000; 2000US-0249264.
17-NOV-2000; 2000US-0249265.
17-NOV-2000; 2000US-0249297.
17-NOV-2000; 2000US-0249299.
17-NOV-2000; 2000US-0249300.
01-DEC-2000; 2000US-0250160.
01-DEC-2000; 2000US-0250391.
05-DEC-2000; 2000US-0251030.
05-DEC-2000; 2000US-0251988.
05-DEC-2000; 2000US-0256719.
06-DEC-2000; 2000US-0251479.
08-DEC-2000; 2000US-0251856.
08-DEC-2000; 2000US-0251868.
08-DEC-2000; 2000US-0251869.
08-DEC-2000; 2000US-0251989.
08-DEC-2000; 2000US-0251990.
11-DEC-2000; 2000US-0254097.
05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-483426/52.
P-PSDB; AAM88986.

Nucleic acids encoding human immune/hematopoietic antigen polypeptides, useful for preventing, diagnosing and/or treating cancers and metastasis -

XX
PS Claim 1; SEQ ID NO 682; 3071pp; Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I) CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic CC activity, and can be used in gene therapy and vaccine production. (I) CC proteins and polynucleotides may be used in the prevention, diagnosis and CC treatment of diseases associated with inappropriate (I) expression. For CC example, they may be used to treat disorders associated with decreased CC expression by rectifying mutations or deletions in a patient's genome CC that affect the activity of (I) by expressing inactive proteins or to CC supplement the patients low production of (I). Additionally, (I) CC polynucleotides may be used to produce the secreted (I), by inserting CC the nucleic acids into a host cell and culturing the cell to express the CC protein. (I) proteins and polynucleotides may be used to prevent, CC diagnose and treat immune/haematopoietic related diseases, especially CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703 CC to AAK87694 represent human immune/haematopoietic antigen genomic CC sequences from the present invention. AAK54952 to AAK54950 and AAM82169 CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 561 BP; 128 A; 115 C; 193 G; 125 T; 0 other;

Alignment Scores:

Pred. No.:	92.7	Length:	561
Score:	7.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	12.50%	Indels:	0
DB:	22	Gaps:	0

US-09-980-043-4 (1-56) x AAK61767 (1-561)

Qy 11 ThrCysGlyGluGlnValGly 17
||||||||||||||||||
Db 28 ACTTGCGGAGAGCAAGTGGGC 8

RESULT 13
ABK33134
ID ABK33134 standard; DNA; 651 BP.
XX
AC ABK33134;
XX
DT 08-MAY-2002 (first entry)
XX
DE DNA encoding novel secreted protein Z910197G6P.
XX
KW Protein secretion; mammalian secreted polypeptide; MSP;
KW gene; ss.
XX
OS Homo sapiens.
XX
PN WO200202621-A2.
XX
PD 10-JAN-2002.
XX
PF 28-JUN-2001; 2001WO-US20658.
XX
PR 30-JUN 2000; 2000US-215446P.
XX
PA (ZYMO) ZYMOGENETICS INC.
XX
PI Sheppard PO, Presnell SR;
XX
DR WPI; 2002-147999/19.
DR P-PSDB; AAU83219.
XX
PT Novel isolated mammalian secreted polypeptide useful in therapeutic and
PT diagnostic methods, to direct secretion of other proteins of interest
PT from host cell, as educational tools, and as laboratory practicum kits
PT -
XX
PS Claim 3; Page 351-352; 397pp; English.

```
XX
CC   The invention describes an isolated mammalian secreted polypeptide (MSP)
CC   (I). (I) is useful to direct the secretion of other proteins of interest
CC   from a host cell, to monitor secretion of proteins, to degenerate
CC   sequences comprising all nucleotide sequences encoding a particular
CC   polypeptide, to screen for cell metabolism effecting receptors, for
CC   identifying new target receptors and drug design, for identifying, for
CC   protein purification, for determining the weight of expressed MSP
CC   polypeptides as a ratio to total protein expressed, for identifying
CC   peptide cleavage sites, for coupling amino and carboxy terminal tags, for
CC   amino acid sequence analysis, for monitoring biological activities of the
CC   protein in vitro and in vivo, and to teach analytical skills and as
CC   reagents for the study of cells, receptors, and other binding molecules.
CC   The polynucleotide is useful for radiation hybrid mapping, and somatic
CC   cell genetic technique developed for constructing high-resolution,
CC   contiguous maps of mammalian chromosomes. Reagents disclosed in the
CC   invention may be used to detect metabolic abnormalities characterised by
CC   over or under production of the protein. This sequence encodes an
CC   mammalian secreted polypeptide, described in the method of the invention.
XX
SQ   Sequence 651 BP; 104 A; 204 C; 218 G; 125 T; 0 other;

Alignment Scores:
Pred. No.:                106             Length:         651
Score:                    7.00            Matches:        7
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              12.50%          Indels:         0
DB:                       24              Gaps:           0

US-09-980-043-4 (1-56) x ABK33134 (1-651)

Qy          3 GluSerGlyAlaProIleCys 9
              |||||||||||||||||||
Db        451 GAGAGTGGGGCACCCATCTGC 471

RESULT 14
AAA67127
ID   AAA67127 standard; DNA; 664 BP.
XX
AC   AAA67127;
XX
DT   31-OCT-2000  (first entry)
XX
DE   Eucalyptus grandis cellulose synthase nucleotide sequence SEQ ID NO:107.
XX
KW   Eucalyptus grandis; pinus radiata; Monterey pine; modification;
KW   plant cell wall; polysaccharide; polysaccharide biosynthetic pathway;
KW   transgenic plant; ds.
XX
OS   Eucalyptus grandis.
XX
PN   WO200022092-A2.
XX
PD   20-APR-2000.
XX
PF   08-OCT-1999;   99WO-NZ00169.
XX
PR   13-OCT-1998;   98US-0170862.
PR   11-AUG-1999;   99US-0148426
XX
PA   (GENE-) GENESIS RES & DEV CORP LTD
PA   (FLET-) FLETCHER CHALLENGE FORESTS LTD
XX
PI   Bloksberg LN;
XX
DR   WPI; 2000-339328/29.
DR   P-PSDB; AAB16320.
XX
PT   New genes encoding proteins involved in a plant polysaccharide
PT   biosynthetic pathway, useful for modulating or altering the
PT   polysaccharide content, composition or structure of the plant -
XX
PS   Claim 1; Page 92; 301pp; English.
XX
CC   The present invention describes isolated polynucleotides (PN) comprising
CC   a sequence selected from one of 835 nucleotide sequences given in
CC   AAA67073 to AAA67907, their (reverse) complements, sequences producing
CC   an Expectation (E) value of 0.01 or less compared to the 835 sequences,
CC   sequences at least 50% identical to them, 200, 100, 40 or 20-mers of the
CC   835 sequences or sequences that are degenerately equivalent or allelic
CC   to the 835 sequences. The polynucleotides are used to modify the
CC   activity of a polypeptide involved in a polysaccharide biosynthetic
CC   pathway in the plant. They are especially used to modulate or alter the
CC   polysaccharide content, composition or structure of the plant. AAB16268
CC   to AAB16340 are proteins encoded by some of the polynucleotide sequence
CC   given in the present invention.
XX
SQ   Sequence 664 BP; 201 A; 136 C; 182 G; 145 T; 0 other;

Alignment Scores:
Pred. No.:                108             Length:         664
Score:                    7.00            Matches:        7
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              12.50%          Indels:         0
DB:                       21              Gaps:           0

US-09-980-043-4 (1-56) x AAA67127 (1-664)

Qy         43 GluIleLysGluGlyArgLys 49
              |||||||||||||||||||
Db        237 GAGATTAAGGAAGGAAGAAAA 257

RESULT 15
AAK52619/c
ID   AAK52619 standard; cDNA; 670 BP.
XX
AC   AAK52619;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human polynucleotide SEQ ID NO 2148.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
KW   nervous system disorder; arthritis; inflammation; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157190-A2.
XX
PD   09-AUG-2001.
XX
PF   05-FEB-2001; 2001WO-US04098.
XX
PR   03-FEB-2000; 2000US-0496914.
PR   27-APR-2000; 2000US-0560875.
PR   20-JUN-2000; 2000US-0598075.
PR   19-JUL-2000; 2000US-0620325.
PR   01-SEP-2000; 2000US-0654936.
PR   15-SEP-2000; 2000US-0663561.
PR   20-OCT-2000; 2000US-0693325.
PR   30-NOV-2000; 2000US-0728422.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT,  Asundi V,  Zhou P,  Xu C,  Cao Y,  Ma Y;
PI   Zhao QA,  Wang D,  Wang J,  Zhang J,  Ren F,  Chen R,  Wang ZW;
PI   Xue AJ,  Yang Y,  Wejhrman T,  Goodrich R;
XX
DR   WPI; 2001-476283/51.
DR   P-PSDB; AAM79486.
XX
PT   Nucleic acids encoding polypeptides with cytokine-like activities,
```

useful in diagnosis and gene therapy -

Claim 1; Page 4517; 6221pp; English.

The invention relates to polynucleotides (AAK51456-AAK53435) and the encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to cytokine, cell proliferation or cell differentiation or which may induce production of other cytokines in other cell populations. The polynucleotides and polypeptides are useful in gene therapy, vaccines or peptide therapy. The polypeptides have various cytokine-like activities, e.g. stem cell growth factor activity, haematopoiesis regulating activity, tissue growth factor activity, immunomodulatory activity and activin/inhibin activity and may be useful in the diagnosis and/or treatment of cancer, leukaemia, nervous system disorders, arthritis and inflammation.
Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the sequence listing were missing at the time of publication.

Sequence 670 BP; 91 A; 232 C; 190 G; 157 T; 0 other;

ignment Scores:

ed. No.:	109	Length:	670
ore:	7.00	Matches:	7
rcent Similarity:	100.00%	Conservative:	0
st Local Similarity:	100.00%	Mismatches:	0
ery Match:	12.50%	Indels:	0
:	22	Gaps:	0

-09-980-043-4 (1-56) x AAK52619 (1-670)

```
         14 GluGlnValGlyHisAspAla 20
            ||||||||||||||||||||
        319 GAGCAGGTCGGGCACGACGCC 299
```

arch completed: October 31, 2003, 23:55:48
b time : 44.2147 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 31, 2003, 20:56:23 ; Search time 602.785 Seconds
(without alignments)
3800.589 Million cell updates/sec

Title: US-09-980-043-4
Perfect score: 56
Sequence: 1 MMESGAPICHTCGEQVGHDA.....KSCFEPEIKEGRKVCLRCGS 56

Scoring table: OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 20454813386 residues

Word size: 1

Total number of hits satisfying chosen parameters: 5773140

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09980043/runat_29102003_163930_22663/app_query.fasta_1.1358
-DB=GenEmbl -QFMT=fastap -SUFFIX=oligo.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09980043_@CGN_1_1_6944_@runat_29102003_163930_22663 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	56	100.0	1010	8	AF197911	AF197911 Populus t
	2	44	78.6	3232	8	AF072131	AF072131 Populus t
	3	10	17.9	219622	9	AC009630	AC009630 Homo sapi
	4	9	16.1	137056	2	AC131248	AC131248 Medicago
	5	8	14.3	1215	8	SOYSTF1A	L28003 Glycine max
	6	8	14.3	2958	8	AF267742	AF267742 Arabidops
	7	8	14.3	3109	8	AY099636	AY099636 Arabidops
c	8	8	14.3	86884	3	AC004366	AC004366 Drosophil
c	9	8	14.3	94301	8	ATF28A21	AL035526 Arabidops
	10	8	14.3	96592	6	AX695560	AX695560 Sequence
	11	8	14.3	99698	8	AC008262	AC008262 Genomic s
c	12	8	14.3	155553	3	AC007802	AC007802 Drosophil
c	13	8	14.3	162711	8	AC137070	AC137070 Genomic s
c	14	8	14.3	163954	2	AC017155	AC017155 Drosophil
	15	8	14.3	164403	9	AC027097	AC027097 Homo sapi
c	16	8	14.3	172185	2	AC068072	AC068072 Homo sapi
c	17	8	14.3	187220	2	AC110199	AC110199 Mus muscu
c	18	8	14.3	199075	8	ATCHRIV49	AL161549 Arabidops
c	19	8	14.3	204917	9	AC008040	AC008040 Homo sapi
c	20	8	14.3	214960	2	AC142361	AC142361 Rattus no
	21	8	14.3	222046	10	AL772291	AL772291 Mouse DNA
c	22	8	14.3	223882	10	AC087113	AC087113 Mus Muscu
c	23	8	14.3	224552	2	AC103271	AC103271 Rattus no
c	24	8	14.3	232385	2	AC133770	AC133770 Rattus no
c	25	8	14.3	242109	2	AC128430	AC128430 Rattus no
	26	8	14.3	242745	2	AC109521	AC109521 Rattus no
	27	8	14.3	247468	2	AC128534	AC128534 Rattus no
c	28	8	14.3	258710	2	AC135823	AC135823 Rattus no
c	29	8	14.3	269255	2	AC106165	AC106165 Rattus no
c	30	8	14.3	291250	3	AE003455	AE003455 Drosophil
	31	8	14.3	292913	2	AC106456	AC106456 Rattus no
	32	8	14.3	347582	3	PFMAL4P1	AL034557 Plasmodiu
c	33	7	12.5	198	6	AX118927	AX118927 Sequence
	34	7	12.5	206	8	AF242635	AF242635 Miscanthu
	35	7	12.5	235	6	BD092304	BD092304 Identific
c	36	7	12.5	319	6	AX389945	AX389945 Sequence
	37	7	12.5	324	11	G10134	G10134 human STS C
	38	7	12.5	353	11	G31903	G31903 MCW146 Chic
	39	7	12.5	386	11	G53714	G53714 SHGC-84877
	40	7	12.5	467	6	AX439162	AX439162 Sequence
	41	7	12.5	550	11	BV035114	BV035114 S212P6608
	42	7	12.5	582	11	G96302	G96302 S209P6229FH
	43	7	12.5	616	11	BV043397	BV043397 S212P6031
	44	7	12.5	651	6	AX380737	AX380737 Sequence
c	45	7	12.5	677	11	BV076114	BV076114 S212P6249

ALIGNMENTS

RESULT 1

```
F197911
OCUS       AF197911                1010 bp    DNA     linear   PLN 02-AUG-2000
EFINITION  Populus tremuloides cellulose synthase (CesA) gene, promotor and
           partial cds.
CCESSION   AF197911
ERSION     AF197911.1  GI:9651405
EYWORDS    .
OURCE      Populus tremuloides (quaking aspen)
 ORGANISM  Populus tremuloides
           Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
           Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
           rosids; eurosids I; Malpighiales; Salicaceae; Populus.
EFERENCE   1  (bases 1 to 1010)
 AUTHORS   Wu,L., Joshi,C.P. and Chiang,V.L.
 TITLE     A xylem-specific cellulose synthase gene from aspen (Populus
           tremuloides) is responsive to mechanical stress
 JOURNAL   Plant J. 22 (6), 495-502 (2000)
 MEDLINE   20344632
  PUBMED   10886769
EFERENCE   2  (bases 1 to 1010)
 AUTHORS   Wu,L., Joshi,C.P. and Chiang,V.L.
 TITLE     Direct Submission
 JOURNAL   Submitted (22-OCT-1999) Forestry, Michigan Tech University, 1400,
           Townsend Drive, Houghton, MI 49931, USA
EATURES             Location/Qualifiers
     source         1. .1010
                    /organism="Populus tremuloides"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:3693"
                    /tissue_type="xylem"
     gene           <1. .>1010
                    /gene="CesA"
     misc_feature   1. .840
                    /gene="CesA"
                    /note="contains promotor and 5'UTR"
     mRNA           <841. .>1010
                    /gene="CesA"
                    /product="cellulose synthase"
     CDS            841. .>1010
                    /gene="CesA"
                    /codon_start=1
                    /product="cellulose synthase"
                    /protein_id="AAF91155.1"
                    /db_xref="GI:9651406"
                    /translation="MMESGAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEI
                    KEGRKVCLRCGS"
ASE COUNT      289 a    212 c    216 g    293 t
RIGIN

lignment Scores:
red. No.:              2.5e-55        Length:          1010
core:                  56.00          Matches:         56
ercent Similarity:     100.00%        Conservative:    0
est Local Similarity:  100.00%        Mismatches:      0
uery Match:            100.00%        Indels:          0
B:                     8              Gaps:            0

3-09-980-043-4 (1-56) x AF197911 (1-1010)

/          1 MetMetGluSerGlyAlaProIleCysHisThrCysGlyGluGlnValGlyHisAspAla 20
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
>        841 ATGATGGAATCTGGGGCTCCTATATGCCATACCTGTGGTGAACAGGTGGGGCATGATGCA 900

/         21 AsnGlyGluLeuPheValAlaCysHisGluCysSerTyrProMetCysLysSerCysPhe 40
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
>        901 AATGGGGAGCTATTTGTGGCTTGCCATGAGTGTAGCTATCCCATGTGCAAGTCTTGTTTC 960

/         41 GluPheGluIleLysGluGlyArgLysValCysLeuArgCysGlySer 56
             ||||||||||||||||||||||||||||||||||||||||||||||||||
>        961 GAGTTTGAAATCAAAGAGGGCCGGAAAGTTTGCTTGCGGTGTGGCTCG 1008

ESULT 2
F072131
```

```
LOCUS       AF072131                3232 bp    mRNA    linear   PLN 26-APR-2001
DEFINITION  Populus tremuloides secondary xylem cellulose synthase (celA) mRNA,
            complete cds.
ACCESSION   AF072131
VERSION     AF072131.1  GI:4115904
KEYWORDS    .
SOURCE      Populus tremuloides (quaking aspen)
  ORGANISM  Populus tremuloides
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Malpighiales; Salicaceae; Populus.
REFERENCE   1  (bases 1 to 3232)
  AUTHORS   Wu,L., Joshi,C.P. and Chiang,V.L.
  TITLE     A xylem-specific cellulose synthase gene from aspen (Populus
            tremuloides) is responsive to mechanical stress
  JOURNAL   Plant J. 22 (6), 495-502 (2000)
  MEDLINE   20344632
   PUBMED   10886769
REFERENCE   2  (bases 1 to 3232)
  AUTHORS   Wu,L., Joshi,C.P. and Chiang,V.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JAN-1998) IWR/Forestry, Michigan Tech University,
            1400, Townsend Drive, Houghton, MI 49931, USA
REFERENCE   3  (bases 1 to 3232)
  AUTHORS   Wu,L., Joshi,C.P. and Chiang,V.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JAN-1999) IWR/Forestry, Michigan Tech University,
            1400, Townsend Drive, Houghton, MI 49931, USA
  REMARK    Sequence update by submitter
COMMENT     On Jan 8, 1999 this sequence version replaced gi:3243277.
FEATURES             Location/Qualifiers
     source          1. .3232
                     /organism="Populus tremuloides"
                     /mol_type="mRNA"
                     /db_xref="taxon:3693"
                     /tissue_type="secondary xylem"
     gene            1. .3232
                     /gene="celA"
     CDS             69. .3065
                     /gene="celA"
                     /codon_start=1
                     /product="secondary xylem cellulose synthase"
                     /protein_id="AAD03417.1"
                     /db_xref="GI:4115905"
                     /translation="MMESGAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEI
                     NEGRKVCLRCGSPYDENLLDDVEKKGSGNQSTMASHLNDSQDVGIHARHISSVSTVDS
                     EMNDEYGNPIWKNRVKSCKDKENKKKKRSPKAETEPAQVPTEQQMEEKPSAEASEPLS
                     IVYPIPRNKLTPYRAVIIMRLVILGLFFHFRITNPVDSAFGLWLTSVICEIWFAFSWV
                     LDQFPKWNPVNRETYIERLSARYEREGEPSQLAGVDFFVSTVDPLKEPPLITANTVLS
                     ILAVDYPVDKVSCYVSDDGAAMLSFESLVETAEFARKWVPFCKKFSIEPRAPEFYFSQ
                     KIDYLKDKVQPSFVKERRAMKRDYEEYKVRVNALVAKAQKTPEEGWTMQDGTPWPGNN
                     TRDHPGHDSGLPWEILGARDIEGNELPRLVYVSREKRPGYQHHKKAGAENALVRVSAV
                     LTNAPYILNVDCDHYVNNSKAVREAMCILMDPQVGRDVCYVQFPQRFDGIDKSDRYAN
                     RNVVFFDVNMKGLDGIQGPVYVGTGCVFNRQALYGYGPPSMPSLRKRKDSSSCFSCCC
                     PSKKKPAQDPAEVYRDAKREDLNAAIFNLTEIDNYDEHERSMLISQLSFEKTFGLSSV
                     FIESTLMENGGVPESANSPPFIKEAIQVIGCGYEEKTEWGKQIGWIYGSVTEDILSGF
                     KMHCRGWRSIYCMPVRPAFKGSAPINLSDRLHQVLRWALGSVEIFFSRHCPLWYGFGG
                     GRLKWLQRLAYINTIVYPFTSLPLIAYCTIPAVCLLTGKFIIPTLSNLASMLFLGLFI
                     SIIVTAVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQGFLKMLAGIDTNFTVTA
                     KAAEDAEFGELYMVKWTTLLIPPTTLLIINMSGCAGFSDALNKGYEAWGPLFGKVFFA
                     FWVILHLYPFLKGLMGRQNRTPTIVVLWSVLLASVFSLVWVKINPFVNKVDNTLVAET
                     CISIDC"
BASE COUNT      895 a    668 c    772 g    897 t
ORIGIN

Alignment Scores:
Pred. No.:              5.18e-41       Length:          3232
Score:                  44.00          Matches:         44
Percent Similarity:     100.00%        Conservative:    0
Best Local Similarity:  100.00%        Mismatches:      0
Query Match:            78.57%         Indels:          0
DB:                     8              Gaps:            0
```

```
US-09-980-043-4 (1-56) x AF072131 (1-3232)

Qy          1 MetMetGluSerGlyAlaProIleCysHisThrCysGlyGluGlnValGlyHisAspAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         69 ATGATGGAATCTGGGGCTCCTATATGCCATACCTGTGGTGAACAGGTGGGGCATGATGCA 128

Qy         21 AsnGlyGluLeuPheValAlaCysHisGluCysSerTyrProMetCysLysSerCysPhe 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        129 AATGGGGAGCTATTTGTGGCTTGCCATGAGTGTAGCTATCCCATGTGCAAGTCTTGTTTC 188

Qy         41 GluPheGluIle 44
              ||||||||||||
Db        189 GAGTTTGAAATC 200

RESULT 3
AC009630
LOCUS       AC009630               219622 bp    DNA     linear   PRI 30-MAR-2002
DEFINITION  Homo sapiens chromosome 8, clone RP11-360L9, complete sequence.
ACCESSION   AC009630
VERSION     AC009630.14  GI:19848415
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 219622)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 8, clone RP11-360L9
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 219622)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,C., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
            Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Depayre,E., Devon,K., Dewar,K.,
            Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
            Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
            Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
            Karatas,A., Lehoczky,J., Lieu,C., Locke,K., Macdonald,P.,
            Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
            Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
            Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
            Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
            Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
            Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-AUG-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 219622)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAR-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4  (bases 1 to 219622)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-MAR-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 30, 2002 this sequence version replaced gi:19683283.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L2093
                Center clone name: 360_L_9
            --------------.
FEATURES             Location/Qualifiers
     source          1. .219622
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8"
                     /clone="RP11-360L9"
                     /clone_lib="RPCI-11 Human Male BAC"
     repeat_region   complement(740. .835)
                     /rpt_family="MIR"
     repeat_region   1626. .1730
                     /rpt_family="GA-rich"
     repeat_region   1995. .2019
                     /rpt_family="GC_rich"
     repeat_region   2306. .2368
                     /rpt_family="GA-rich"
     repeat_region   complement(3606. .3702)
                     /rpt_family="L2"
     repeat_region   3987. .4054
                     /rpt_family="MIR3"
     repeat_region   5149. .5504
                     /rpt_family="AluSx"
     repeat_region   5900. .6166
                     /rpt_family="AluSq"
     repeat_region   6260. .6432
                     /rpt_family="MIR"
     repeat_region   6482. .6863
                     /rpt_family="MLT1A1"
     repeat_region   6877. .7052
                     /rpt_family="MIR"
     repeat_region   7370. .7671
                     /rpt_family="AluSx"
```

```
repeat_region   7817. .7914
                /rpt_family="MER5A"
repeat_region   8016. .8126
                /rpt_family="MSTB1"
repeat_region   8117. .8182
                /rpt_family="MSTB1"
repeat_region   8183. .8489
                /rpt_family="AluSx"
repeat_region   8490. .8555
                /rpt_family="MSTB1"
repeat_region   9370. .9677
                /rpt_family="AluSx"
repeat_region   complement(9719. .9832)
                /rpt_family="MIR"
repeat_region   complement(9928. .10135)
                /rpt_family="MIR"
repeat_region   complement(10239. .10348)
                /rpt_family="L2"
repeat_region   complement(10751. .11044)
                /rpt_family="AluJb"
repeat_region   complement(11087. .11322)
                /rpt_family="L1MA7"
repeat_region   11324. .11411
                /rpt_family="AluY"
repeat_region   11412. .11695
                /rpt_family="AluSc"
repeat_region   complement(11696. .12441)
                /rpt_family="L1MA7"
repeat_region   complement(12442. .12744)
                /rpt_family="AluSx"
repeat_region   complement(12745. .12803)
                /rpt_family="L1MA7"
repeat_region   complement(12804. .13115)
                /rpt_family="AluSx"
repeat_region   complement(13116. .13499)
                /rpt_family="L1MA7"
repeat_region   13500. .13792
                /rpt_family="AluSx"
repeat_region   complement(13793. .14061)
                /rpt_family="L1MA7"
repeat_region   complement(14093. .14239)
                /rpt_family="MIR3"
repeat_region   14535. .14557
                /rpt_family="(GGGGA)n"
repeat_region   complement(14642. .14785)
                /rpt_family="L1PA8"
repeat_region   15668. .15702
                /rpt_family="G-rich"
repeat_region   complement(16451. .16575)
                /rpt_family="MER41A"
repeat_region   16576. .16871
                /rpt_family="AluSx"
repeat_region   complement(16872. .17310)
                /rpt_family="MER41A"
repeat_region   19693. .19988
                /rpt_family="AluJo"
repeat_region   19992. .20023
                /rpt_family="Alu"
repeat_region   20024. .20288
                /rpt_family="AluSg/x"
repeat_region   20289. .20300
                /rpt_family="Alu"
repeat_region   21314. .22905
                /rpt_family="SVA"
repeat_region   24489. .24786
                /rpt_family="AluSq"
repeat_region   complement(25042. .25134)
                /rpt_family="MIR3"
repeat_region   complement(26856. .26938)
                /rpt_family="MIR"

lignment Scores:
ed. No.:        4.11        Length:        219622

Score:                  10.         Matches:        10
Percent Similarity:     100.00%     Conservative:   0
Best Local Similarity:  100.00%     Mismatches:     0
Query Match:            17.86%      Indels:         0
DB:                     9           Gaps:           0

US-09-980-043-4 (1-56) x AC[illegible] (1-219622)

Qy          45 LysGluGlyArgLysValCysLeuArgCys 54
               ||||||||||||| ||| |||||| |||
Db      124899 AAGGAAGGCAGAAAAGTGTGCTTACGCTGT 124928

RESULT 4
AC131248
LOCUS       AC131248              137056 bp    DNA     linear   HTG 21-MAY-2003
DEFINITION  Medicago truncatula clone mth2 28n22, WORKING DRAFT SEQUENCE, 3
            ordered pieces
ACCESSION   AC131248
VERSION     AC131248.5  GI:[illegible]
KEYWORDS    HTG; HTGS_PHASE[illegible]
SOURCE      Medicago truncatula [illegible]
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 137056)
  AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
  TITLE     Medicago truncatula BAC Clone mth2 28n22
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 137056)
  AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-AUG-2002) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   3  (bases 1 to 137056)
  AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAY-2003) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
COMMENT     On Apr 23, 2003 this sequence version replaced gi:29893694.
            -------------- Genome Center
            Center: Department Of Chemistry And Biochemistry
            The University Of Oklahoma
            Center code:UOKNOR
            ---------- ------- --
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *        1     2327: contig of 2327 bp in length
            *     2328     2427: gap of unknown length
            *     2428    39732: contig of 37305 bp in length
            *    39733    39832: gap of unknown length
            *    39833   137056: contig of 97224 bp in length.
FEATURES            Location/Qualifiers
     source         1. .137056
                    /organism="Medicago truncatula"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:3880"
                    /clone="mth2-28n22"
                    /clone_lib="Medicago truncatula BAC library H2"
BASE COUNT    45991 a  21335 c  22570 g  46956 t    204 others
```

```
ORIGIN

Alignment Scores:
Pred. No.:                39.9          Length:         137056
Score:                    9.00          Matches:        9
Percent Similarity:       100.00%       Conservative:   0
Best Local Similarity:    100.00%       Mismatches:     0
Query Match:              16.07%        Indels:         0
DB:                       2             Gaps:           0

US-09-980-043-4 (1-56) x AC131248 (1-137056)

Qy         21 AsnGlyGluLeuPheValAlaCysHis 29
              |||||||||||||||||||||||||||
Db      70245 AATGGAGAGCTTTTTGTGGCTTGTCAT 70271

RESULT 5
SOYSTF1A
LOCUS       SOYSTF1A                1215 bp    mRNA    linear   PLN 05-MAR-1998
DEFINITION  Glycine max TGACG-motif binding protein (STF1) mRNA, complete cds.
ACCESSION   L28003
VERSION     L28003.1  GI:2934883
KEYWORDS    .
SOURCE      Glycine max (soybean)
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 1215)
  AUTHORS   Cheong,Y.H., Yoo,C.M., Park,J.M., Ryu,G.R., Goekjian,V.H.,
            Nagao,R.T., Key,J.L., Cho,M.J. and Hong,J.C.
  TITLE     STF1 is a novel TGACG-binding factor with a zinc-finger motif and a
            bZIP domain which heterodimerizes with GBF proteins
  JOURNAL   Plant J. 15 (2), 199-209 (1998)
  MEDLINE   98388650
   PUBMED   9721678
REFERENCE   2  (bases 1 to 1215)
  AUTHORS   Hong,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-SEP-1995) Biochemistry, Gyeongsang National
            University, Gwaza dong 900, Chinju, KyungNam 660-701, Korea
COMMENT     On Mar 5, 1998 this sequence version replaced gi:986964.
FEATURES             Location/Qualifiers
     source          1. .1215
                     /organism="Glycine max"
                     /mol_type="mRNA"
                     /strain="Williams"
                     /db_xref="taxon:3847"
                     /tissue_type="hypocotyl"
                     /dev_stage="seedling"
     gene            1. .1215
                     /gene="STF1"
     CDS             4. .984
                     /codon_start=1
                     /product="TGACG-motif binding factor"
                     /protein_id="AAC05017.1"
                     /db_xref="GI:2934884"
                     /translation="MERSGGMVTGSHERNELVRVRHGSDSGSKPLKNLNGQICQICGD
                     TIGLTATGDLFVACHECGFPLCHSCYEYELKNVSQSCPQCKTTFTSRQEGAEVEGDDD
                     DEDDADDLDNGINYGQGNNSKSGMLWEEDADLSSSSGHDSHIPNPHLVNGQPMSGEFP
                     CATSDAQSMQTTSDPMGQSEKVHSLPYADPKQPGPESDEEIRRVPEIGGESAGTSASR
                     PDAGSNAGTERAQGTGDSQKKRGRSPADKESKRLKRLLRNRVSAQQARERKKAYLIDL
                     ETRVKDLEKKNSELKERLSTLQNENQMLRQILKNTTASRRGSNSGTNNAV"
     misc_feature    118. .255
                     /gene="STF1"
                     /note="encodes zinc-finger domain"
     misc_feature    751. .813
                     /gene="STF1"
                     /note="encodes basic DNA-binding region (BR)"
     misc_feature    829. .915
                     /gene="STF1"
                     /note="encodes leucine zipper region (ZR)"
     polyA_site      1215
                     /gene="STF1"
BASE COUNT      395 a    211 c    310 g    299 t
ORIGIN

Alignment Scores:
Pred. No.:                11.1          Length:         1215
Score:                    8.00          Matches:        8
Percent Similarity:       100.00%       Conservative:   0
Best Local Similarity:    100.00%       Mismatches:     0
Query Match:              14.29%        Indels:         0
DB:                       8             Gaps:           0

US-09-980-043-4 (1-56) x SOYSTF1A (1-1215)

Qy         24 LeuPheValAlaCysHisGluCys 31
              |||||||||||||||||||||||
Db        163 CTCTTTGTTGCTTGTCATGAGTGT 186

RESULT 6
AF267742
LOCUS       AF267742                2958 bp    mRNA    linear   PLN 16-FEB-2001
DEFINITION  Arabidopsis thaliana cellulose synthase catalytic subunit (IRX1)
            mRNA, complete cds.
ACCESSION   AF267742
VERSION     AF267742.1  GI:12836996
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 2958)
  AUTHORS   Taylor,N.G., Laurie,S. and Turner,S.R.
  TITLE     Multiple cellulose synthase catalytic subunits are required for
            cellulose synthesis in Arabidopsis
  JOURNAL   Plant Cell 12 (12), 2529-2540 (2000)
  MEDLINE   21066939
   PUBMED   11148295
REFERENCE   2  (bases 1 to 2958)
  AUTHORS   Taylor,N.G. and Turner,S.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-MAY-2000) Biological Sciences, University of
            Manchester, 3.614 Stopford Building, Oxford Road, Manchester M13
            9PT, UK
FEATURES             Location/Qualifiers
     source          1. .2958
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
     gene            1. .2958
                     /gene="IRX1"
     CDS             1. .2958
                     /gene="IRX1"
                     /codon_start=1
                     /product="cellulose synthase catalytic subunit"
                     /protein_id="AAK08700.1"
                     /db_xref="GI:12836997"
                     /translation="MMESRSPICNTCGEEIGAKSNGEFFVACHECSFPICKACLEYEF
                     KEGRRICLRCGNPYDENVFDDVETKTSKTQSIVPTQTNNTSQDSGIHARHISTVSTID
                     SELNDEYGNPIWKNRVESWKDKKDKKSKKKKKDPKATKAEQQDAQVPTQQHMEDMPPN
                     TESGATDVLSVVIPIPRTKITSYRIVIIMRLIILALFFNYRITHPVDSAYGLWLTSVI
                     CEIWFAVSWVLDQFPKWSPINRETYIDRLSARFEREGEQSQLAAVDFFVSTVDPLKEP
                     PLITANTVLSILALDYPVDKVSCYVSDDGAAMLSFESLVETADFARKWVPFCKKYSIE
                     PRAPEFYFSLKIDYLRDKVQPSFVKERRAMKRDYEEFKIRMNALVAKAQKTPEEGWTM
                     QDGTSWPGNNTRDHPGMIQVFLGYSGARDIEGNELPRLVYVSREKRPGYQHHKKAGAE
                     NALVRVSAVLTNAPFILNLDCDHYVNNSKAVREAMCFLMDPVVGQDVCFVQFPQRFDG
                     IDKSDRYANRNIVFFDVNMRGLDGIQGPVYVGTGTVFRRQALYGYSPPSKPRILPQSS
                     SSSCCCLTKKKQPQDPAEIYKDAKREELDAAIFNLGDLDNYDEYDRSMLISQTSFEKT
                     FGLSTVFIESTLMENGGVPDSVNPSTLIKEAIHVISCGYEEKTEWGKEIGWIYGSITE
                     DILTGFKMHCRGWRSIYCMPLRPAFKGSAPINLSDRLHQVLRWALGSVEIFLSRHCPL
                     WYGCSGGRLKLLQRLAYINTIVYPFTSLPLVAYCTLPAICLLTGKFIIPTLSNLASML
                     FLGLFISIILTSVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQGFLKMLAGLDT
```

NFTVTSKTADDLEFGELYIVKWTTLLIPPTSLLIINLVGVVAGFSDALNKGYEAWGPL
FGKVFFAFWVILHLYPFLKGLMGRQNRTPTIVILWSILLASVFSLVWVRINPFVSKTD
TTSLSLNCLLIDC"

SE COUNT 818 a 610 c 658 g 872 t
IGIN

ignment Scores:

ed. No.:	23.3	Length:	2958
ore:	8.00	Matches:	8
rcent Similarity:	100.00%	Conservative:	0
st Local Similarity:	100.00%	Mismatches:	0
ery Match:	14.29%	Indels:	0
:	8	Gaps:	0

-09-980-043-4 (1-56) x AF267742 (1-2958)

25 PheValAlaCysHisGluCysSer 32
||||||||||||||||||||||||
73 TTTGTGGCTTGTCATGAGTGTAGT 96

SULT 7
099636
CUS AY099636 3109 bp mRNA linear PLN 06-MAY-2002
FINITION Arabidopsis thaliana cellulose synthase-like protein (At4g18780) mRNA, complete cds.
CESSION AY099636
RSION AY099636.1 GI:20466339
YWORDS FLI_CDNA.
URCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
FERENCE 1 (bases 1 to 3109)
AUTHORS Nguyen,M., Karlin-Neumann,G., Southwick,A., Lam,B., Miranda,M., Palm,C.J., Bowser,L., Jones,T., Banh,J., Carninci,P., Chen,H., Cheuk,R., Chung,M.K., Hayashizaki,Y., Ishida,J., Kamiya,A., Kawai,J., Kim,C., Lin,J., Liu,S.X., Narusaka,M., Pham,P.K., Sakano,H., Sakurai,T., Satou,M., Seki,M., Shinn,P., Yamada,K., Shinozaki,K., Ecker,J., Theologis,A. and Davis,R.W.
TITLE Direct Submission
JOURNAL Submitted (24-APR-2002) DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA
MMENT e-mail for correspondence: arab@sequence.stanford.edu

RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN Arabidopsis Full-Length cDNA'): Seki,M., Narusaka,M., Ishida,J., Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J., Hayashizaki,Y. and Shinozaki,K.

The Salk, Stanford, PGEC (SSP) Consortium members carried out the sequencing and annotation of the RAFL cDNAs: Nguyen,M., Southwick,A., Karlin-Neumann,G., Lam,B., Miranda,M., Palm,C.J., Bowser,L., Jones,T., Banh,J., Chen,H., Cheuk,R., Chung,M.K., Kim,C., Lin,J., Liu,S.X., Pham,P.K., Sakano,H., Shinn,P., Yamada,K., Ecker,J., Theologis,A. and Davis,R.W.

Nguyen,M, (SSP/Stanford) and Seki,M. (RIKEN GSC) contributed equally to this work. Shinozaki,K. (RIKEN GSC) and Davis,R.W. (SSP/Stanford) contributed equally to this work as PIs.

ATURES	Location/Qualifiers
source	1. .3109
	/organism="Arabidopsis thaliana"
	/mol_type="mRNA"
	/db_xref="taxon:3702"
	/chromosome="4"
	/clone="RAFL09-65-M12 (R24904)"
	/note="This clone is in pBluescript
	ecotype: Columbia"
gene	1. .3109
	/gene="At4g18780"
	/note="synonym: F28A21.19"
CDS	68. .3025
	/gene="At4g18780"
	/codon_start=1
	/product="cellulose synthase-like protein"
	/protein_id="AAM20487.1"
	/db_xref="GI:20466340"
	/translation="MMESRSPICNTCGEEIGVKSNGEFFVACHECSFPICKACLEYEF
	KEGRRICLRCGNPYDENVFDDVETKTSKTQSIVPTQTNNTSQDSGIHARHISTVSTID
	SELNDEYGNPIWKNRVESWKDKKDKKSKKKKKDPKATKAEQHEAQIPTQQHMEDTPPN
	TESGATDVLSVVIPIPRTKITSYRIVIIMRLIILALFFNYRITHPVDSAYGLWLTSVI
	CEIWFAVSWVLDQFPKWSPINRETYIDRLSARFEREGEQSQLAAVDFFVSTVDPLKEP
	PLITANTVLSILALDYPVDKVSCYVSDDGAAMLSFESLVETADFARKWVPFCKKYSIE
	PRAPEFYFSLKIDYLRDKVQPSFVKERRAMKRDYEEFKIRMNALVAKAQKTPEEGWTM
	QDGTSWPGNNTRDHPGMIQVFLGYSGARDIEGNELPRLVYVSREKRPGYQHHKKAGAE
	NALVRVSAVLTNAPFILNLDCDHYVNNSKAVREAMCFLMDPVVGQDVCFVQFPQRFDG
	IDKSDRYANRNIVFFDVNMRGLDGIQGPVYVGTGTVFRRQALYGYSPPSKPRILPQSS
	SSSCCCLFKKQPQDPSEIYKPAKREELDAAIFNLGDLDNYDEYDRSMLISQTSFEKT
	FGLSTVFIESTLMENGGVPESVNPATLIKEAIHVISCGYEEKTEWGKEIGWIYGSITE
	DILTGFKMHCRGWRSIYCMPKRPAFKGSAPINLSDRLHQVLRWALGSVEIFLSRHCPL
	WYGCSGGRLKLLQRLAYINIIVYPFTSIPLVAYCTLPAICLLTGKFIIPTLSNLASML
	FLGLFISIILTSVLELRWSGVSIEDLWRNEQFWVIGGVSAHLFAVFQGFLKMLAGLDT
	NFTVTSKTADDLEFGELYIVKWTTLLIPPTSLLIINLVGVVAGFSDALNKGYEAWGPL
	FGKVFFAFWVILHLYPFLKGLMGRQNRTPTIVILWSILLASVFSLVWVRINPFVSKTD
	TTSLSLNCLLIDC"

BASE COUNT 859 a 631 c 687 g 932 t
ORIGIN

Alignment Scores:

Pred. No.:	24.3	Length:	3109
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	14.29%	Indels:	0
DB:	8	Gaps:	0

US-09-980-043-4 (1-56) x AY099636 (1-3109)

Qy 25 PheValAlaCysHisGluCysSer 32
||||||||||||||||||||||||
Db 140 TTTGTGGCTTGTCATGAGTGTAGT 163

RESULT 8
AC004366/c
LOCUS AC004366 86884 bp DNA linear INV 02-APR-1998
DEFINITION Drosophila melanogaster (P1 DS07550 (D60)) DNA sequence, complete sequence.
ACCESSION AC004366 AC002884 AC000792 AC000789 AC000788 AC000793 AC000787 AC000795 L81536 L81535 L81534 L81533 L81532 L81531 L81530 L81529 L81528 L81527 L81526 L81525 L81524 L81523 L81522 L81521 AC000790 AC000794 L81520 AC000791
VERSION AC004366.1 GI:3006206
KEYWORDS HTG.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 86884)
AUTHORS Celniker,S.E., Aghavani,A., Arcaina,T.T., Baxter,E., Blazej,R.G., Chew,M., Doyle,C.M., Farfan,D.E., Flanagan,J., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L., Kim,S.H., Lee,B., Lomotan,M.A., Mazda,P., Mok,M.S., Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Punch,D., Snir,E., Twomey,B., Wan,K.H., Whitelaw,K.R., Yee,A., Zhang,R., Zieran,L.L. and Kimmel,B.
TITLE Sequencing of Drosophila chromosome 2R, region 58A3-58A4
JOURNAL Unpublished (1997)
REFERENCE 2 (bases 1 to 86884)
AUTHORS Celniker,S.E., George,R.A., Galle,R., Svirskas,R.R., Hoskins,R.A., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G., Chavez,C., Chew,M., Doyle,C.M., Farfan,D.E., Flanagan,J., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L., Kim,S.H., Lee,B.,

Lomotan,M.A., Mak,J., Mazda,P., Mok,M.S., Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Punch,E., Snir,E., Twomey,B., Wan,K.H., Whitelaw,K.R., Yee,A., Zhang,R., Zieran,L.L. and Kimmel,B.E.
TITLE Direct Submission
JOURNAL Submitted (07-MAR-1998) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT On Apr 2, 1998 this sequence version replaced gi:2943870.
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive Web site
(http://fruitfly.berkeley.edu/sequence-archive.html) or send email to drosophila@genome.lbl.gov.
Library location: 62-79.
This P1 was assembled from the following subclones: 1_a11 (AC002884), 2_c5 (AC000792), 2_c11, 1_g3 (AC000789), 1_e8 (AC000788), 2_d12 (AC000793), 1_b9 (AC000787), 2_h5 (AC000795), 1_d5 (L81536), 2_h11 (L81535), 1_a9 (L81534), 1_g1 (L81533), 2_h6 (L81532), 2_e2 (L81531), 2_b1 (L81530), 1_a1 (L81529), 2_g8 (L81528), 2_b7 (L81527), 2_g4 (L81526), 2_f8 (L81525), 1_e10 (L81524), 1_e6 (L81523), 2_h3 (L81522), 1_c12 (L81521), 1_g4 (AC000790), 1_b10, 2_f5 (AC000794), 2_h12 (L81520), 2_b2 (AC000791), 1_h7.
FEATURES Location/Qualifiers
source 1. .86884
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/chromosome="2R"
/map="58A3-58A4"
/clone="P1 DS07550 (D60)"
BASE COUNT 24688 a 17939 c 17839 g 26418 t
ORIGIN

Alignment Scores:

Pred. No.:	392	Length:	86884
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	14.29%	Indels:	0
DB:	3	Gaps:	0

US-09-980-043-4 (1-56) x AC004366 (1-86884)

Qy 46 GluGlyArgLysValCysLeuArg 56
||||||||||||||||||||||||
Db 64560 GAAGGAAGGAAGGTTTGTTTAAGA 64567

RESULT 9
ATF28A21/c
LOCUS ATF28A21 94301 bp DNA linear PLN 08-SEP-1999
DEFINITION Arabidopsis thaliana DNA chromosome 4, BAC clone F28A21 (ESSA project).
ACCESSION AL035526
VERSION AL035526.1 GI:4539378
KEYWORDS .
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1
AUTHORS Bevan,M., Mueller,M.W., Muendlein,A., Felber,R., Bancroft,I., Mewes,H.W., Mayer,K.F.X., Lemcke,K. and Schueller,C.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 94301)
AUTHORS EU Arabidopsis sequencing project.
TITLE Direct Submission
JOURNAL Submitted (08-SEP-1999) MIPS, at the Max Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail: schuelle@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de Project Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK, E-mail: michael.bevan@bbsrc.ac.uk
COMMENT Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/.
FEATURES Location/Qualifiers
source 1. .94301
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/variety="Columbia"
/db_xref="taxon:3702"
/chromosome="4"
gene 9652. .10227
/gene="F28A21.20"
CDS 9652. .10227
/gene="F28A21.20"
/note="similarity to predicted proteins, Arabidopsis thaliana
contains EST gb:T46643, R29928"
/codon_start=1
/product="putative protein"
/protein_id="CAB37446.1"
/db_xref="GI:4539380"
/translation="MSSDRHTPTKDPPDHPSSSSNHHKQPLPPQPQQPLSRYESQKRR
DWNTFVQYLKSQNPPLMMSQFDYTHVLSFLRYLDQFGKTKVHHQACVFFGQPDPPGPC
TCPLKQAWGSLDALIGRLRAAYEEHGGGSPDTNPFANGSIRVHLREVRESQAKARGIP
YRKKKRRKTKNEVVVVKKDVANSSTPNQSFT"
exon 9652. .10227
/gene="F28A21.20"
/number=1
gene 13217. .13711
/gene="F28A21.30"
CDS 13217. .13711
/gene="F28A21.30"
/note="similarity to various predicted proteins, Arabidopsis thaliana"
/codon_start=1
/product="putative protein"
/protein_id="CAB37447.1"
/db_xref="GI:4539381"
/translation="MESSKQKRCRSSVVETIEAPLPLVWSILRSFDKPQAYQRFVKSC
TMRSGGGGGKGGEGKGSVRDVTLVSGFPADFSTERLEELDDESHVMVVSIIGGNHRLV
NYKSKTKVVASPEDMAKKTVVVESYVVDVPEGTSEEDTIFFVDNIIRYNLTSLAKLTK
KMMK"
exon 13217. .13711
/gene="F28A21.30"
/number=1
gene 16036. .17454
/gene="F28A21.40"
CDS 16036. .17454
/gene="F28A21.40"
/note="similarity to predicted protein, Arabidopsis thaliana"
/codon_start=1
/product="putative protein"
/protein_id="CAB37448.1"
/db_xref="GI:4539382"
/translation="MEGKKLNLYAPLPSIRRIPSMIERSSELENKKTTITRPELRSCP
STKQETPVFVLPDQSFDHLTEPASIPFMWEQIPGKPKDDMATLIQESGLLETDDEEED
EVDEDLDTVSSNVSFSVNCSTSGVSEIEKTGERSDCDEKSRESLDLMMSRFLPAAKAM
ALQTHQKHQSSYNSSEQKLITQNREALVARQRSQLVAEHEHFAIVQSLYDDLNIDDDT
EDDENDDDGDDYGHVDHKIYPAEVTKKACGFLPRLCAKNSFKFLNPVPLDSVSNRSMK
HSGDQTSLPNWSTRRLSGPISPYRTSCSDSGFLGTPEKPESPKRLNRGISKSQELYPT
RTRREVLPNYSNSGEIKMPRNSISTPSRIQRTTIHDSRFLVEEVNRRRNNKSGNLQQT
SSEYTPALSPPPLPETPSRSWLGRTLLPPVNPRPYGVVLGQVGIKKLNQEVLESTKWE
TIVKTSYVHNDHARYSQVLITL"
exon 16036. .17454
/gene="F28A21.40"
/number=1
gene 19501. .22435

```
                     /gene="F28A21.50"
     CDS             join(<19501. .19652,19742. .19813,19902. .20045,20144. .20287,
                     20375. .20469,20559. .20987,21071. .21261,21342. .21474,
                     21531. .21709,21893. .22084,22178. .22435)
                     /gene="F28A21.50"
                     /codon_start=1
                     /product="putative protein (fragment)"
                     /protein_id="CAB37449.1"
                     /db_xref="GI:4539383"
                     /translation="FVLNSGSALLKFRARVNSDPHGTLANWNVSGINDLCYWSGVTCV
                     DGKVQILDLSGYSLEGTLAPELSQLSDLRSLILSRNHFSGGIPKEYGSFENLEVLDLR
                     ENDLSGQIPPELSNGLSLKHLLLSGNKFSDDMRIKIVRLQSSYEVRLKKSPKLSPLAV
                     LGCINRKLGHCVSRNRIIQVKKVEAIVFRIKATSRRFLKAFPSFLEETDIYKRRELLE
                     ETSNLAAEPAPSAPSPSPGIITEASPRSSGSFPAVTNAKKRRPPLVPPVPSPDKGSTS
                     PDISKNQPQDNKQSKGSKHVWLYVVIAVASFVGLLIIVAVIFFCRKRAVKSIGPWKTG
                     LSGQLQKAFVTGVPKLNRSELETACEDFSNIIETFDGYTVYKGTLSSGVEIAVASTAI
                     AESKEWTRAMEMAYRRKIDTLSRINHKNFVNLIGYCEEDDPFNRMMVFEYAPNGTLFE
                     HLHVNNVPRLSDKETEHLDWSARMRIIMGTAYCLQHMHGMNPPMAHTDFNSSEIYLTD
                     DYAAKVSEIPFNLEARLNPKKHVSGDLEQTSLILPPEPEANVHSFGVLMLEIINGKLS
                     FSDEYGSIEQWASKYLEKDDLGEMIDPSLKTFKEEELEVICDVIRECLKTEQRQRPSM
                     KDVAEQLKQVINITPEKATPRSSPLWWAELEILSSEAT"
     exon            19501. .19652
                     /gene="F28A21.50"
                     /number=1
     intron          19653. .19741
                     /gene="F28A21.50"
                     /number=1
     exon            19742. .19813
                     /gene="F28A21.50"
                     /number=2
     intron          19814. .19901
                     /gene="F28A21.50"
                     /number=2
     exon            19902. .20045
                     /gene="F28A21.50"
                     /number=3
     intron          20046. .20143
                     /gene="F28A21.50"
                     /number=3
     exon            20144. .20287
                     /gene="F28A21.50"
                     /number=4
     intron          20288. .20374
                     /gene="F28A21.50"
                     /number=4
     exon            20375. .20469
                     /gene="F28A21.50"
                     /number=5
     intron          20470. .20558
                     /gene="F28A21.50"
                     /number=5
     exon            20559. .20987
                     /gene="F28A21.50"
                     /number=6
     intron          20988. .21070
                     /gene="F28A21.50"
                     /number=6
     exon            21071. .21261
                     /gene="F28A21.50"
                     /number=7
     intron          21262. .21341
                     /gene="F28A21.50"
                     /number=7
     exon            21342. .21474
                     /gene="F28A21.50"
                     /number=8
     intron          21475. .21530
                     /gene="F28A21.50"
                     /number=8
     exon            21531. .21709
                     /gene="F28A21.50"
                     /number=9
     intron          21710. .21892
                     /gene="F28A21.50"
                     /number=9
     exon            21893. .22084
                     /gene="F28A21.50"
                     /number=10
     intron          22085. .22177
                     /gene="F28A21.50"
                     /number=10
     exon            22178. .22435
                     /gene="F28A21.50"
                     /number=11
     gene            22901. .24271
                     /gene="F28A21.60"
     gene            complement(join(22901. .23176,23858. .24271))
                     /gene="F28A21.60"
     CDS             complement(join(22901. .23176,23858. .24271))
                     /gene="F28A21.60"
                     /note="similarity to transcription factor CBF3.1, Zea
                     mays. [U27837]"
                     /codon_start=1
                     /product="putative protein"
                     /protein_id="CAB37450.1"
                     /db_xref="GI:4539384"
                     /translation="MRNLVEEKFLEFYESWVIQLELYLHQLLIAHNNNTMSETELRHL
                     ISKLTTHHKAYYTAKWAAIREDVLAFFGSVWLNPLENACSWLTGWKPSMVFRMVDRLR
                     KSRVVLVEAQVKKLEELRVKTKFDFQKIERFMERYQVAMADRKMVELARLGCHVGGES
                     VMVVEAAVRGLSMGLEKMVKAADCVRLKTLKGILDILTPPQCVEFIAAAATFQVQLRR
                     WGNRRHYVTHS"
     exon            complement(22901. .23176)
                     /gene="F28A21.60"
                     /number=1
     intron          complement(23177. .23857)
                     /gene="F28A21.60"
                     /number=1
     exon            complement(23858. .24271)

Alignment Scores:
Pred. No.:                420             Length:          94301
Score:                    8.00            Matches:         8
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              14.29%          Indels:          0
DB:                       8               Gaps:            0

US-09-980-043-4 (1-56) x ATF28A21 (1-94301)

Qy         25 PheValAlaCysHisGluCysSer 32
              ||||||||||||||||||||||||
Db      75505 TTTGTGGCTTGTCATGAGTGTAGT 75482

RESULT 10
AX695560
LOCUS       AX695560               96592 bp    DNA     linear   PAT 31-MAR-2003
DEFINITION  Sequence 1187 from Patent WO03008583.
ACCESSION   AX695560
VERSION     AX695560.1  GI:29418712
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus
REFERENCE   1
  AUTHORS   Morris,D.W. and Engelhard,E.K.
  TITLE     Novel compositions and methods for cancer
  JOURNAL   Patent: WO 03008583-A 1187 30 JAN-2003;
            Sagres Discovery (US)
FEATURES             Location/Qualifiers
     source          1. .96592
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
BASE COUNT    27935 a  19437 c  18902 g  29171 t   1147 others
ORIGIN
```

```
Alignment Scores:
Pred. No.:              428             Length:         96592
Score:                  8.00            Matches:        8
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            14.29%          Indels:         0
DB:                     6               Gaps:           0

US-09-980-043-4 (1-56) x AX695560 (1-96592)

Qy          2 MetGluSerGlyAlaProIleCys 9
              ||||||||||||||||||||||||
Db      89327 ATGGAAAGTGGGGCCCCGATTTGC 89350

RESULT 11
AC008262
LOCUS       AC008262               99698 bp    DNA     linear   PLN 11-OCT-2000
DEFINITION  Genomic sequence for Arabidopsis thaliana BAC F4N2 from chromosome
            I, complete sequence.
ACCESSION   AC008262
VERSION     AC008262.4  GI:5757471
KEYWORDS    HTG.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 99698)
  AUTHORS   Shinn,P., Brooks,S., Buehler,E., Chao,Q., Dunn,P., Khan,S., Kim,C.,
            Altafi,H., Araujo,R., Conn,L., Conway,A.B., Gonzalez,A.,
            Hansen,N.F., Huizar,L., Kremenetskaia,I., Lenz,C., Li,J., Liu,S.,
            Luros,S., Rowley,D., Schwartz,J., Toriumi,M., Vysotskaia,V., Yu,G.,
            Davis,R.W., Federspiel,N.A., Theologis,A. and Ecker,J.R.
  TITLE     Genomic sequence for Arabidopsis thaliana BAC F4N2 from chromosome
            I
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 99698)
  AUTHORS   Ecker,J.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-JUL-1999) Arabidopsis thaliana Genome Center,
            Department of Biology, University of Pennsylvania, 38th Street and
            Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
REFERENCE   3  (bases 1 to 99698)
  AUTHORS   Ecker,J.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-AUG-1999) Arabidopsis thaliana Genome Center,
            Department of Biology, University of Pennsylvania, 38th Street and
            Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
REFERENCE   4  (bases 1 to 99698)
  AUTHORS   Chao,Q., Brooks,S., Buehler,E., Johnson-Hopson,C., Khan,S., Kim,C.,
            Shinn,P., Altafi,H., Bei,B., Chin,C., Chiou,J., Choi,E., Conn,L.,
            Conway,A., Gonzalez,A., Hansen,N., Howing,B., Koo,T., Lam,B.,
            Lee,J., Lenz,C., Li,J., Liu,A., Liu,J., Liu,S., Mukharsky,N.,
            Nguyen,M., Palm,C., Pham,P., Sakano,H., Schwartz,J., Southwick,A.,
            Thaveri,A., Toriumi,M., Vaysberg,M., Yu,G., Davis,R.,
            Federspiel,N., Theologis,A. and Ecker,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-JAN-2000) Arabidopsis thaliana Genome Center,
            Department of Biology, University of Pennsylvania, 38th and
            Hamilton Walk, Philadelphia, PA 19104-6018, USA
REFERENCE   5  (bases 1 to 99698)
  AUTHORS   Cheuk,R., Shinn,P., Brooks,S., Buehler,E., Chao,Q.,
            Johnson-Hopson,C., Khan,S., Kim,C., Altafi,H., Bei,B., Chin,C.,
            Chiou,J., Choi,E., Conn,L., Conway,A., Gonzalez,A., Hansen,N.,
            Howing,B., Koo,T., Lam,B., Lee,J., Lenz,C., Li,J., Liu,A., Liu,J.,
            Liu,S., Mukharsky,N., Nguyen,M., Palm,C., Pham,P., Sakano,H.,
            Schwartz,J., Southwick,A., Thaveri,A., Toriumi,M., Vaysberg,M.,
            Yu,G., Davis,R., Federspiel,N., Theologis,A. and Ecker,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JUN-2000) Arabidopsis thaliana Genome Center,
            Department of Biology, University of Pennsylvania, 38th and
            Hamilton Walk, Philadelphia, PA 19104-6018, USA
REFERENCE   6  (bases 1 to 99698)
  AUTHORS   Chao,Q., Brooks,S., Buehler,E., Johnson-Hopson,C., Khan,S., Kim,C.,
            Shinn,P., Altafi,H., Bei,B., Chin,C., Chiou,J., Choi,E., Conn,L.,
            Conway,A., Gonzalez,A., Hansen,N., Howing,B., Koo,T., Lam,B.,
            Lee,J., Lenz,C., Li,J., Liu,A., Liu,J., Liu,S., Mukharsky,N.,
            Nguyen,M., Palm,C., Pham,P., Sakano,H., Schwartz,J., Southwick,A.,
            Thaveri,A., Toriumi,M., Vaysberg,M., Yu,G., Davis,R.,
            Federspiel,N., Theologis,A. and Ecker,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-OCT-2000) Arabidopsis thaliana Genome Center,
            Department of Biology, University of Pennsylvania, 38th and
            Hamilton Walk, Philadelphia, PA 19104-6018, USA
COMMENT     On Aug 21, 1999 this sequence version replaced gi:5733851.
FEATURES             Location/Qualifiers
     source          1. .99698
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3702"
                     /chromosome="1"
                     /clone="F4N2"
     CDS             complement(join(467. .577,593. .1117))
                     /note="unknown protein; similar to EST gb|AI996556.1"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="F4N2.1"
                     /protein_id="AAF27074.1"
                     /db_xref="GI:6730653"
                     /translation="MEDLWKRAKSFAEEAAKKSQTITLQSSSTTFVNLVTETAKKSKE
                     FALEASKKADSINVSEFVAETAKKSKEFAAEVSTKADQLKVVALKQADQIQNIKSIAD
                     IIPGGLGGSGSGGGGSEISESELVSFGITDDLREFVKGLTSATFQAFPEQGIVVLLLL
                     LLYLCFGIEDLKCVMMNRRKFLIWKQVRRMFGRIFLSGRRDMPLLFLVLSR"
     CDS             join(10069. .10115,10212. .10257,10355. .10469,10562. .10629,
                     10712. .11314,11395. .11499,11581. .11934)
                     /note="putative uridylyl transferase gb|AAD20075; similar
                     to ESTs gb|AA713247, gb|AA597876"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="F4N2.2"
                     /protein_id="AAF27052.1"
                     /db_xref="GI:6730631"
                     /translation="MDNEYEKLIRRMNPPRVVIDNDSCKKATVIRVDSANEYGILLEV
                     VQILTDLNLTITKAYISSDGGWFMDVFNVTDQDGNKVTDEVVLDYIQKSLGPEACFST
                     SMRSVGVIPSTDSTVIELTGCDRPGLLSELSAVLTHLKCSVLNAEIWTHNTRAAAVMQ
                     VTDDLTGCGISDPERLSRIKNLLRNVLKGSNTPREAKTVVSHGEVHTDRRLHQMMFED
                     RDYEHRLVDDDSSIQDERQRPDVCVDNWLDKDYSVVTVRCKDRPKLLFDTVCTLTDMQ
                     YVVFHGSVDTEGTEAFQEYYVRHIDGSPVKSEAEKQRVIQCLEAAIKRRVSEGLKLEL
                     CTTDRVGLLSNVTRIFRENSLTVTRAEVKTKGGKALNTFYVSDASGYSIDAKTIDSIR
                     QTIGQTILKVKNNPQEQQQRQKSPSHESPTRFLFGGLFKSKSFVNFGLVRSYS"
     CDS             join(13830. .14004,16439. .16484,16984. .18391,18549. .18662,
                     18749. .18898,18982. .19128)
                     /note="hypothetical protein"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="F4N2.3"
                     /protein_id="AAF27053.1"
                     /db_xref="GI:6730632"
                     /translation="MEDSGRKGDPFPGFEMRVHNDDCVFRRSLSCRLQRRAPCPLLLP
                     PRPPPSAEAPSTTTVFKPPSSSSSVVVVERMPTKRCSANGIMEEFILVDDDDVIILDF
                     PESSKRKAPGTSRTRKESIMPRVISIDDDDDETENVQKAGSSSSGNLQTSAEVDNDDC
                     QFIQEKCATFRFSKCKQTFSTMPSHGIRFGLGSDTYSDFSESDCSDCEILEGSHRDVR
                     EQWEQAALLRKMKKAGKAGLSEEAGPSNLHCDTNFEPGFERRTEQHDQTSFFFAARNP
                     DGGKAGLTEEAGPSNLHCDNNFRPGFESRTEQHDQTSFFFTASNPDGEKETSSTFFGA
                     DGRIPDMTTYFGSVMEEDNQAKHTANSTFSSKEDFRGPSPLSPGIRVEHERFDTPPTH
                     FTSPKEPIHQTSFKKVVEQPNKVQVQCETSLSPEENQNKNVAYQSYTKEVEEEEDASK
                     LSPPQTSNGYGDTTHVFGTQSNGARCEHGISLDESREPIIDPIPSTSGQVQGTNGTAP
                     AIDVMLNREILKETDEYKRAQEEEWESRQRQLQIQADEAQKQRKRRKLENMRKLEMER
                     RQKERVEEVRETQKKDEENMNMKEKVRAEITKSLKLLELKSFNMAALLRGLGIQVGGG
                     ISPLPNEVHAAYKRAVLKFHPDRASRGGDIKQQVEAEEKFKLIARMKDKFKLIAP"
     CDS             complement(join(19276. .19546,19647. .19925,20017. .20156,
                     20283. .20336,20490. .20683,20780. .20948,21047. .21742,
                     21900. .22028,22154. .22276,22365. .22424,22517. .22569,
                     22660. .22822,22900. .23244))
                     /note="hypothetical protein"
```

```
                     /codon_start=1
                     /evidence=not_experimental
                     /product="F4N2.4"
                     /protein_id="AAF27073.1"
                     /db_xref="GI:6730652"
                     /translation="MGKDNKMKKKPNKKGERRMGPDAVAMKAKTQKVDNPFESIRSRR
                     KFDILGKKRKGEERFVSVSRTRAVDKRKNTLEKEYEQSLKSSVFLDKRIGEQNDELGE
                     FDKGIIRSQRQRQLKLAKKSMYNLSDGEEDVYEDGALGGSSVKDDFDSGLLSDEDLQD
                     DDLEASASKRLKHLNRNREVDASGEEERRKSKKEVMEEIIMKSKLGRMEKAKQKEEKG
                     KLMDELDKNFKSLVNSEAMESLTKPFVAEENTRDPYLLSLNDMSMEIRARPSERTKTP
                     EEIAQKEREKLEALEEERKKRMQETEELSDGDEEIGGEESTKRLTVISGDDLGDSFSV
                     EEDKPKRGWIDDVLEREDNVDNSESDEDEDSESEEEEDDDGESDGGDEKQRKGHHLED
                     WEQSDDELGAELEDEEEDDDEEDDDEEDAELRVHKKLKNDYAAPYKGEGLSGTVKEKT
                     NMKKMSSTQRDIPFMIDPPKNFEELLALVEDCSNEDVILIVNRIRIAHSIKIKAENRK
                     KMQVKFNPALKGLILYFAVLASKKPLNFDLLNMLVKPLIEMSMEIPYFAAICARQRLL
                     KTRSQFCEAIKNPEDGCWPSLKTLFLLRLWSLIFPCSDFRHAVMTPSILLMCEYLMRC
                     PISSGRDIAIGSFLCSIVLLTKSHQHLRNLRYAIEQDLFYHFMELKSLTPLLCIQDNV
                     KEVMPLNPLKIMNEPADSPYFSSDDFRASILSSVVETLEGFVEINGGLSSFPEIFMPI
                     STLLHQIGNQEKIPQTLKEKLEDVAKLIEKKTDDHHKERKPLSMRKHKPVAIRMVNPK
                     FEENFVPGMDNDPDKYRSDLKKLKRKLKREARGAVRELRKDSYFMSTVKAKEKAAHEQ
                     ERAEKHGKAWAFLQEQEHAFKSGQLGKGKGKKRRR"
CDS                  complement(join(23987. .24280,24633. .24748,24837. .25098))
                     /note="unknown protein; similar to ESTs gb|N38009,
                     gb|AI997072.1"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="F4N2.5"
                     /protein_id="AAF27072.1"
                     /db_xref="GI:6730651"
                     /translation="MGKKGEGPFVSRLRANVRVQPPTTTSQHGCSHEPPTTSIEIKGR
                     RIIVVVDSCSEAKNALLWTLSHCAQPQDSILLLHFLKAKTSQSGDLANKEEGEDESCD
                     KPTTSRADKKVSALKTMCELKRPEVKTEVVFVKGDEKGPTIVKEARENEASLLVLGQK
                     KQHATWRLLMVWASQARPVTKHDFVEYCINNSPCMAIAVRKRGKKLGGYTLTTKRHKD
                     FWLLA"
CDS                  28229. .28462
                     /note="similar to FMRFamide precursor isolog gi|1871179"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="F4N2.6"
                     /protein_id="AAF27054.1"
                     /db_xref="GI:6730633"
                     /translation="MILTFWRENTILRFWRENTILRFWRENTILRFWRENTILRFWRE
                     NTILRFWRKNTILRFWRKNTILRFWWKNMILRF"
CDS                  complement(29356. .30648)
                     /note="hypothetical protein"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="F4N2.7"
                     /protein_id="AAF27071.1"
                     /db_xref="GI:6730650"
                     /translation="MNFLSKSEKKKKLISSPTLESLTDKIIRKMASPTLALAQSPPPK
                     SPAVSVSQRNPHCWSKLPLDLMQLVFERLAFLDFERAKSVCSSWQFGSKQSKPNNQIP
                     WMILFPTDKNYCLLFNPEDKEKLYKTQHLGDDFAKSIVLATYRSWLLMQPRYEELEDQ
                     TLDQEFHLYIKDLLTCERINLPAFESDIFGLSHPILWIDDKTKDYLVIGTINRETMVS
                     FKNGDNSWKKFPELPKSSCTDMCLNMIYKDHKLHYLDYSNLYIYDFFGEFPREAFRIS
                     VREFVGYATNPYGYDEFPEVPLKLKLNRYIYNMIVTVRGDVLIVASLHFSMAETWEFI
                     ICKMDSSKVNKWEEIVSLGDESILLGLGITVLAKDMEGITCNSIYFTADDFYEDYEEN
                     EIFIYNLDTNKVEIPHQFVSSSIPSAHARWFLPTFKRD"
CDS                  join(31830. .32157,32241. .32740,32828. .32956,33042. .33118,
                     33198. .33291)
                     /note="similar to aspartic proteinase cDNA gi|3157937"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="F4N2.8"
                     /protein_id="AAF27055.1"
                     /db_xref="GI:6730634"
                     /translation="MSPSTTTKMQILLLFLVIQLCHGSNDTVSISLKRHTLNVGGTSF
                     GGLKNFDGVVFYGEISVGSPPQKFNVVFDTGSTDLWVPSKEWPEETDHKHPKFDKDAS
                     KTCRLMKGGEVNIAYETGSVVGILAQDNVNVGGVVIKSQDLFLARNPDTYFRSVKFDG
                     VIGLGIKSSRAQGSVTVWENMVKQKLITKPIFSLYLRPHKGDGGEDPNGGQIMFGGFD
                     PKQFKGEHVYVPMKLSDDRWKIKMSKIYINGKPAINFCDDVECTAMVDSGSTDIFGPD

lignment Scores:

Pred. No.:                440                 Length:          99698
Score:                    8.00                Matches:         8
Percent Similarity:       100.00%             Conservative:    0
Best Local Similarity:    100.00%             Mismatches:      0
Query Match:              14.29%              Indels:          0
DB:                       8                   Gaps:            0

US-09-980-043-4 (1-56) x AC008262 (1-99698)

Qy          45 LysGluGlyArgLysValCysLeu 62
                  ||||||||||||||:|||:||||
Db       40218 AAAGAAGGCAGGAAAGTTTGTTTA 40241

RESULT 12
AC007802/c
LOCUS       AC007802               155553 bp    DNA     linear   INV 27-APR-2001
DEFINITION  Drosophila melanogaster, chromosome 2R, region 58A1-58A2, BAC clone
            BACR07I11, complete sequence.
ACCESSION   AC007802
VERSION     AC007802.7  GI:7208705
KEYWORDS    HTG.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 155553)
  AUTHORS   Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
            Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,
            Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
            Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,
            Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
            Ferriera,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
            Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
            Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,
            McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
            Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
            Phouanenavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
            Stapleton,M., Strong,R., Svirskas,R., Tector,C., Williams,S.M.,
            Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
  TITLE     Sequencing of Drosophila chromosome 2R, region 58A1-58A2
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 155553)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
            Rubin,G.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUN-1999) Drosophila Genome Center, Lawrence Berkeley
            Laboratory, MS 64 121, Berkeley, CA 94720, USA
COMMENT     On Apr 27, 2001 this sequence version replaced gi:7208705.
            Sequence submitted by:
            Berkeley Drosophila Genome Project
            Lawrence Berkeley National Laboratory, MS 64-121
            Berkeley, CA 94720
            This sequence was assembled using end sequences from a whole genome
            shotgun and from subclones of this BAC and its neighboring clones.
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www.fruitfly.org/sequence/) or send email
            to bdgp@fruitfly.berkeley.edu.
FEATURES                Location/Qualifiers
     source             1. .155553
                        /organism="Drosophila melanogaster"
                        /mol_type="genomic DNA"
                        /strain="y; cn bw sp"
                        /db_xref="taxon:7227"
                        /chromosome="2R"
```

```
                     /map="58A1-58A2"
                     /clone="BACR07I11 (D648)"
                     /clone_lib="RPCI-98 (Roswell Park Cancer Institute
                     Drosophila melanogaster BAC library, partial EcoRI in
                     pBACe3.6)"
BASE COUNT    45766 a  31823 c  31935 g  46029 t
ORIGIN

Alignment Scores:
Pred. No.:                638           Length:        155553
Score:                    8.00          Matches:       8
Percent Similarity:       100.00%       Conservative:  0
Best Local Similarity:    100.00%       Mismatches:    0
Query Match:              14.29%        Indels:        0
DB:                       3             Gaps:          0

US-09-980-043-4 (1-56) x AC007802 (1-155553)

Qy          46 GluGlyArgLysValCysLeuArg 53
               ||||||||||||||||||||||||
Db      145607 GAAGGAAGGAAGGTTTGTTTAAGA 145584

RESULT 13
AC137070/c
LOCUS       AC137070              162711 bp    DNA     linear   PLN 31-DEC-2002
DEFINITION  Genomic sequence for Oryza sativa, Nipponbare strain, clone
            OSJNBa0035J18, from chromsome 3, complete sequence.
ACCESSION   AC137070
VERSION     AC137070.2  GI:27436772
KEYWORDS    HTG.
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 162711)
  AUTHORS   McCombie,W.R., de la Bastide,M., Spiegel,L., Preston,R.,
            Nascimento,L., Zutavern,T., Balija,V., Bell,M., Miller,B.,
            Katzenberger,F., Muller,S., Yang,C., Dike,S., O'Shaughnessy,A.,
            Palmer,L. and Dedhia,N.
  TITLE     Genomic sequence for Oryza sativa, Nipponbare strain, clone
            OSJNBa0035J18, from chromosome 3, complete sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 162711)
  AUTHORS   McCombie,W.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-2002) Lita Annenberg Hazen Genome Center, Cold
            Spring Harbor Laboratories, 1, Bungtown Road, Cold Spring Harbor,
            NY 11724, USA
REFERENCE   3  (bases 1 to 162711)
  AUTHORS   McCombie,W.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-DEC-2002) Lita Annenberg Hazen Genome Center, Cold
            Spring Harbor Laboratories, 1, Bungtown Road, Cold Spring Harbor,
            NY 11724, USA
COMMENT     On Dec 31, 2002 this sequence version replaced gi:25046309.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest.  The nucleotide
            sequence of this BAC clone was generated by combining Syngenta and
            Cold Spring Harbor Laboratory Genome Center sequencing data.
            Clone OSJNBa0035J18 overlaps clone OSJNBb0014I10 (AC126222) from
            base 1 to base 11914.  The overlap is from base 1 to base 11914
            (complemented) on clone OSJNBb0014I10.
FEATURES             Location/Qualifiers
     source          1. .162711
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="genomic DNA"
                     /strain="Nipponbare"
                     /db_xref="taxon:39947"
                     /chromosome="3"
                     /clone="OSJNBa0035J18"
                     /clone_lib="HindIII"
                     /note="(japonica cultivar-group)"
BASE COUNT    44584 a  35530 c  35591 g  47006 t
ORIGIN

Alignment Scores:
Pred. No.:                662           Length:        162711
Score:                    8.00          Matches:       8
Percent Similarity:       100.00%       Conservative:  0
Best Local Similarity:    100.00%       Mismatches:    0
Query Match:              14.29%        Indels:        0
DB:                       8             Gaps:          0

US-09-980-043-4 (1-56) x AC137070 (1-162711)

Qy          44 IleLysGluGlyArgLysValCys 51
               |||||||||||||||||||||||
Db      41613 ATCAAAGAGGGGCGAAAAGTTTGT 41590

RESULT 14
AC017155/c
LOCUS       AC017155              163954 bp    DNA     linear   HTG 09-DEC-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***.
ACCESSION   AC017155
VERSION     AC017155.1  GI:6553831
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 163954)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as COM:10209684 by the submitter.
            For more information on this record e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .163954
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
BASE COUNT    47211 a  34064 c  34178 g  48501 t
ORIGIN

Alignment Scores:
Pred. No.:                666           Length:        163954
Score:                    8.00          Matches:       8
Percent Similarity:       100.00%       Conservative:  0
Best Local Similarity:    100.00%       Mismatches:    0
Query Match:              14.29%        Indels:        0
DB:                       2             Gaps:          0

US-09-980-043-4 (1-56) x AC017155 (1-163954)

Qy          46 GluGlyArgLysValCysLeuArg 53
               ||||||||||||||||||||||||
Db      35598 GAAGGAAGGAAGGTTTGTTTAAGA 35575

RESULT 15
AC027097
LOCUS       AC027097              164403 bp    DNA     linear   PRI 01-SEP-2002
DEFINITION  Homo sapiens chromosome 18, clone RP11-35G9, complete sequence.
ACCESSION   AC027097
VERSION     AC027097.10  GI:22597588
```

```
EYWORDS    HTG.
OURCE      Homo sapiens (human)
 ORGANISM  Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
EFERENCE   1  (bases 1 to 164403)
 AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
 TITLE     Homo sapiens chromosome 18, clone RP11-35G9
 JOURNAL   Unpublished
EFERENCE   2  (bases 1 to 164403)
 AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
           Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
           Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
           Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
           Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
           Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
           Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
           Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
           Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
           Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
           Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
           McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
           Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
           Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
           O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
           Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
           Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
           Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
           Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
           Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
           Young,G., Zainoun,J., Zimmer,A. and Zody,M.
 TITLE     Direct Submission
 JOURNAL   Submitted (26-MAR-2000) Whitehead Institute/MIT Center for Genome
           Research, 320 Charles Street, Cambridge, MA 02141, USA
EFERENCE   3  (bases 1 to 164403)
 AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
           Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
           Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
           Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
           Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
           Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
           Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
           Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
           Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
           McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
           Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
           O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
           Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
           Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
           Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
           Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
           Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
           Zembek,L., Zimmer,A. and Zody,M.
 TITLE     Direct Submission
 JOURNAL   Submitted (26-AUG-2002) Whitehead Institute/MIT Center for Genome
           Research, 320 Charles Street, Cambridge, MA 02141, USA
EFERENCE   4  (bases 1 to 164403)
 AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
           Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
           Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
           Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
           Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
           Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
           Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
           Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
           Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
           McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
           Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
           O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
           Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
           Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
           Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
           Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
           Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
           Zembek,L., Zimmer,A. and Zody,M.
 TITLE     Direct Submission
 JOURNAL   Submitted (01-SEP-2002) Whitehead Institute/MIT Center for Genome
           Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT    On Sep 1, 2002 this sequence version replaced gi:22477078.
           All repeats were identified using RepeatMasker:
           Smit, A.F.A. & Green, P. (1996-1997)
           http://ftp.genome.washington.edu/RM/RepeatMasker.html
           ---------- Genome Center
               Center: Whitehead Institute/ MIT Center for Genome Research
               Center code: WIBR
               Web site: http://www-seq.wi.mit.edu
               Contact: sequence_submissions@genome.wi.mit.edu
           ----------   Project Information
               Center project name: L8741
               Center clone name: 35 G 9
           ---------- Genome Center
FEATURES             Location/Qualifiers
     source          1..164403
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="18"
                     /map="18"
                     /clone="RP11-35G9"
                     /clone_lib="RPCI-11 Human Male BAC"
     repeat_region   complement(625..918)
                     /rpt_family="AluJo"
     unsure          645..652
                     /note="<30 qual SNGL region."
     unsure          1093..1098
                     /note="<30 qual SNGL region."
     repeat_region   1301..1601
                     /rpt_family="AluSx"
     repeat_region   complement(1767..2229)
                     /rpt_family="MLT1E"
     repeat_region   2235..2525
                     /rpt_family="AluSg"
     repeat_region   2595..2911
                     /rpt_family="AluJo"
     repeat_region   complement(2936..3001)
                     /rpt_family="MLT1E"
     repeat_region   3217..3394
                     /rpt_family="AluSp"
     repeat_region   complement(3399..3508)
                     /rpt_family="MER58A"
     repeat_region   complement(3546..3610)
                     /rpt_family="L2"
     repeat_region   complement(3958..4058)
                     /rpt_family="L2"
     repeat_region   4173..4475
                     /rpt_family="AluSq"
     repeat_region   complement(4477..4622)
                     /rpt_family="MLT1B"
     repeat_region   complement(4682..4802)
                     /rpt_family="L2"
     repeat_region   complement(4812..5128)
                     /rpt_family="AluSq"
     repeat_region   5757..5805
                     /rpt_family="L1MB4"
     repeat_region   5806..6081
                     /rpt_family="AluJo"
     repeat_region   6222..6444
                     /rpt_family="AluJb"
     repeat_region   6456..6508
                     /rpt_family="L1MB4"
     repeat_region   6549..6727
                     /rpt_family="L1"
     repeat_region   6728..7016
                     /rpt_family="AluJo"
     repeat_region   7017..7295
                     /rpt_family="L1"
```

```
     repeat_region   7297. .7676
                     /rpt_family="L1PA8"
     repeat_region   7754. .7779
                     /rpt_family="AT_rich"
     repeat_region   complement(7872. .8648)
                     /rpt_family="LTR17"
     repeat_region   8857. .8991
                     /rpt_family="L1M"
     repeat_region   9065. .9282
                     /rpt_family="AluSg/x"
     repeat_region   complement(9399. .9557)
                     /rpt_family="L1ME"
     repeat_region   complement(9741. .10045)
                     /rpt_family="AluY"
     repeat_region   10067. .10230
                     /rpt_family="MIR"
     repeat_region   10567. .10615
                     /rpt_family="MER112"
     repeat_region   10616. .10909
                     /rpt_family="AluSx"
     repeat_region   10910. .11076
                     /rpt_family="MER112"
     repeat_region   11094. .11392
                     /rpt_family="(TA)n"
     repeat_region   12095. .12161
                     /rpt_family="(TA)n"
     repeat_region   complement(12162. .12447)
                     /rpt_family="AluSb"
     repeat_region   12609. .12635
                     /rpt_family="(TG)n"
     repeat_region   12635. .12673
                     /rpt_family="(TA)n"
     repeat_region   complement(12720 .12797)
                     /rpt_family="MER58"
     repeat_region   13621. .13932
                     /rpt_family="AluSq"
     repeat_region   14451. .14748
                     /rpt_family="AluSx"
     repeat_region   14958. .15037
                     /rpt_family="CT-rich"
     repeat_region   complement(15662. .15961)
                     /rpt_family="AluSg"
     repeat_region   16365. .16391
                     /rpt_family="(TTTA)n"
     repeat_region   complement(16392 .16672)
                     /rpt_family="AluSq"
     repeat_region   complement(16893 [illegible]
                     /rpt_family="[illegible]
     repeat_region   complement([illegible]
                     /rpt_family "[illegible]

Alignment Scores:
Pred. No.:              668             Length:         164403
Score:                  8.00            Matches:        8
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            14.29%          Indels:         0
DB:                     9               Gaps:           0

US-09-980-043-4 (1-56) x AC027097 (1-164403)

Qy          45 LysGluGlyArgLysValCysLeu 52
               ||||||||||||||||||||||||||
Db       65903 AAGGAAGGAAGGAAGGTGTGTTTA 65926


Search completed: November  1, 2003, [illegible]:41
Job time : 683.785 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 29, 2003, 18:34:34 ; Search time 2.97872 Seconds
(without alignments)
3217.546 Million cell updates/sec

Title: US-09-980-043-4
Perfect score: 56
Sequence: 1 MMESGAPICHTCGEQVGHDA.....KSCPEPEIKEGRKVCLRCGS 56

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 642050 seqs, 171146064 residues

Word size : 0

Total number of hits satisfying chosen parameters: 642050

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	7	12.5	44	10	US-09-881-752A-170	Sequence 170, App
2	7	12.5	216	10	US-09-893-737-290	Sequence 290, App
3	7	12.5	974	10	US-09-838-539-6	Sequence 6, Appli
4	6	10.7	8	14	US-10-029-756-7	Sequence 7, Appli
5	6	10.7	40	9	US-09-250-883-19	Sequence 19, Appl
6	6	10.7	101	14	US-10-004-381-12	Sequence 12, Appl
7	6	10.7	157	12	US-10-029-386-34258	Sequence 34258, A
8	6	10.7	193	14	US-10-001-289-47	Sequence 47, Appl
9	6	10.7	202	9	US-09-815-242-5521	Sequence 5521, Ap
10	6	10.7	202	9	US-09-815-242-12448	Sequence 12448, A
11	6	10.7	202	11	US-09-966-521-30	Sequence 30, Appl
12	6	10.7	202	12	US-10-429-094-30	Sequence 30, Appl
13	6	10.7	308	10	US-09-738-626-3592	Sequence 3592, Ap
14	6	10.7	322	9	US-09-726-643-72	Sequence 72, Appl
15	6	10.7	322	14	US-10-042-141-72	Sequence 72, Appl
16	6	10.7	329	9	US-09-925-301-904	Sequence 904, App
17	6	10.7	340	9	US-09-250-883-17	Sequence 17, Appl
18	6	10.7	346	15	US-10-043-487-217	Sequence 217, App
19	6	10.7	349	15	US-10-146-772-126	Sequence 126, App
20	6	10.7	359	12	US-10-340-779A-12	Sequence 12, Appl
21	6	10.7	359	14	US-10-029-756-2	Sequence 2, Appli
22	6	10.7	428	12	US-10-012-762-19	Sequence 19, Appl
23	6	10.7	452	11	US-09-798-051-9	Sequence 9, Appli
24	6	10.7	452	12	US-09-798-053-9	Sequence 9, Appli
25	6	10.7	453	9	US-09-726-643-46	Sequence 46, Appl
26	6	10.7	453	14	US-10-042-141-46	Sequence 46, Appl
27	6	10.7	457	10	US-09-373-967-6	Sequence 6, Appli
28	6	10.7	465	9	US-09-884-260A-52	Sequence 52, Appl
29	6	10.7	469	15	US-10-128-714-3516	Sequence 3516, Ap
30	6	10.7	489	9	US-09-726-643-82	Sequence 82, Appl
31	6	10.7	489	10	US-09-925-300-1151	Sequence 1151, Ap
32	6	10.7	489	14	US-10-042-141-82	Sequence 82, Appl
33	6	10.7	528	15	US-10-128-714-8516	Sequence 8516, Ap
34	6	10.7	646	9	US-09-777-710A-1	Sequence 1, Appli
35	6	10.7	651	9	US-09-815-242-5331	Sequence 5331, Ap
36	6	10.7	651	9	US-09-815-242-12682	Sequence 12682, A
37	6	10.7	670	9	US-09-777-710A-15	Sequence 15, Appl
38	6	10.7	678	15	US-10-156-761-7585	Sequence 7585, Ap
39	6	10.7	739	10	US-09-957-635-2	Sequence 2, Appli
40	6	10.7	1054	9	US-09-798-042-87	Sequence 87, Appl
41	6	10.7	1084	15	US-10-229-193-8	Sequence 8, Appli
42	6	10.7	1131	10	US-09-738-626-5762	Sequence 5762, Ap
43	6	10.7	1135	11	US-09-759-130B-42	Sequence 42, Appl
44	6	10.7	1272	15	US-10-118-513A-2	Sequence 2, Appli
45	6	10.7	1272	15	US-10-118-513A-8	Sequence 8, Appli

ALIGNMENTS

RESULT 1
US-09-881-752A-170
; Sequence 170, Application US/09881752A
; Patent No. US20020115078A1
; GENERAL INFORMATION:
; APPLICANT: Kleanthous, Harold
; APPLICANT: Al-Garawi, Amal
; APPLICANT: Miller, Charles
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Oomen, Raymond P.
; TITLE OF INVENTION: Identification of Polynucleotides
; TITLE OF INVENTION: Encoding No. US20020115078A1el Helicobacter Polypeptides in t
; TITLE OF INVENTION: Genome
; FILE REFERENCE: 06132/041002
; CURRENT APPLICATION NUMBER: US/09/881,752A
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 08/833,457
; PRIOR FILING DATE: 1997-04-01
; NUMBER OF SEQ ID NOS: 370
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 170
; LENGTH: 44
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-881-752A-170

Query Match 12.5%; Score 7; DB 10; Length 44;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 24 LFVACHE 30
|||||||
Db 13 LFVACHE 19

RESULT 2
US-09-893-737-290
; Sequence 290, Application US/09893737

```
Patent No. US20020110855A1
GENERAL INFORMATION:
 APPLICANT: Sheppard, Paul O.
 APPLICANT:  Presnell, Scott R.
 TITLE OF INVENTION: MAMMALIAN SECRETED PROTEINS
 FILE REFERENCE: 00-41
 CURRENT APPLICATION NUMBER: US/09/893,737
 CURRENT FILING DATE:  2001-06-28
 PRIOR APPLICATION NUMBER: US 60/215,446
 PRIOR FILING DATE: 2000-06-30
 NUMBER OF SEQ ID NOS: 329
 SOFTWARE: FastSEQ for Windows Version 3.0
SEQ ID NO 290
  LENGTH: 216
  TYPE: PRT
  ORGANISM: Homo sapiens
-09-893-737-290

Query Match             12.5%;  Score 7;  DB 10;  Length 216;
Best Local Similarity   100.0%;  Pred. No. 8.7;
Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

          3 ESGAPIC 9
            |||||||
        151 ESGAPIC 157


SULT 3
-09-838-539-6
Sequence 6, Application US/09838539
Patent No. US20020129401A1
GENERAL INFORMATION:
 APPLICANT: Stalker, D. et al.
 TITLE OF INVENTION: Plant Cellulose Synthase and Promoter
 TITLE OF INVENTION:  Sequences
 FILE REFERENCE: 15621/03/US
 CURRENT APPLICATION NUMBER: US/09/838,539
 CURRENT FILING DATE:  2001-04-18
 PRIOR APPLICATION NUMBER: 60/029,987
 PRIOR FILING DATE: 1996-10-29
 PRIOR APPLICATION NUMBER: 08/960,048
 PRIOR FILING DATE: 1997-10-29
 NUMBER OF SEQ ID NOS: 12
 SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 6
  LENGTH: 974
  TYPE: PRT
  ORGANISM: Gossypim hirsutum
-09-838-539-6

Query Match             12.5%;  Score 7;  DB 10;  Length 974;
Best Local Similarity   100.0%;  Pred. No. 32;
Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

         25 FVACHEC 31
            |||||||
         25 FVACHEC 31


SULT 4
-10-029-756-7
Sequence 7, Application US/10029756
Publication No. US20020108147A1
  GENERAL INFORMATION:
       APPLICANT: Thomas, Terry L.
       TITLE OF INVENTION: PRODUCTION OF GAMMA LINOLENIC ACID BY A
                           DELTA 6-DESATURASE
       NUMBER OF SEQUENCES: 27
       CORRESPONDENCE ADDRESS:
            ADDRESSEE: Scully, Scott, Murphy & Presser
            STREET: 400 Garden City Plaza
            CITY: Garden City
            STATE: New York
            COUNTRY: United States
            ZIP: 11530
       COMPUTER READABLE FORM:
            MEDIUM TYPE: Floppy disk
            COMPUTER: IBM PC compatible
            OPERATING SYSTEM: PC-DOS/MS-DOS
            SOFTWARE: PatentIn Release #1.0, Version #1.25
       CURRENT APPLICATION DATA:
            APPLICATION NUMBER: US/10/029,756
            FILING DATE: 21-Dec-2001
            CLASSIFICATION: <Unknown>
       PRIOR APPLICATION DATA:
            APPLICATION NUMBER: 08/934,254
            FILING DATE: <Unknown>
       ATTORNEY/AGENT INFORMATION:
            NAME: Presser, Leopold
            REGISTRATION NUMBER: 19,827
            REFERENCE/DOCKET NUMBER: 81[illegible]
       TELECOMMUNICATION INFORMATION:
            TELEPHONE [illegible]
            TELEFAX: (516) [illegible]
            TELEX: 230 [illegible]
  INFORMATION FOR SEQ ID NO: 7:
       SEQUENCE CHARACTERISTICS:
            LENGTH: 8 amino acids
            TYPE: amino acid
            TOPOLOGY: linear
       MOLECULE TYPE: Peptide
       SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-10-029-756-7

  Query Match             10.7%;  Score 6;  DB 14;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 5.8e+05;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          16 VGHDAN 21
               ||||||
Db           2 VGHDAN 7


RESULT 5
US-09-250-883-19
; Sequence 19, Application US/09250883
; Patent No. US20020042049A1
;  GENERAL INFORMATION:
;    APPLICANT:  Russell, John
;    APPLICANT:  Colpitts, Tracey
;    TITLE OF INVENTION:  REAGENTS AND METHODS USEFUL
;    TITLE OF INVENTION:  FOR DETECTING DISEASE OF THE BREAST
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Abbott Laboratories
;      STREET:  100 Abbott Park Road
;      CITY:  Abbott Park
;      STATE:  IL
;      COUNTRY:  USA
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/250,883
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/889,316
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Becker, Cheryl L
;      REGISTRATION NUMBER:  35,441
;      REFERENCE/DOCKET NUMBER:  6131.US.01
```

```
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  847/935-1729
;      TELEFAX:  847/938-2623
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  19:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  40 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  No. US20020042049A1e
US-09-250-883-19

  Query Match             10.7%;  Score 6;  DB 9;  Length 40;
  Best Local Similarity   100.0%;  Pred. No. 24;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          49 KVCLRC 54
               ||||||
Db          15 KVCLRC 20


RESULT 6
US-10-004-381-12
; Sequence 12, Application US/10004381
; Publication No. US20020155578A1
; GENERAL INFORMATION:
;  APPLICANT: SZOSTAK,  JACK W.
;  APPLICANT:  WILSON, DAVID S.
;  APPLICANT:  KEEFE, ANTHONY D.
;  TITLE OF INVENTION: STREPTAVIDIN-BINDING PEPTIDES AND USES
;  TITLE OF INVENTION:  THEREOF
;  FILE REFERENCE: 00786/388002
;  CURRENT APPLICATION NUMBER: US/10/004,381
;  CURRENT FILING DATE:  2001-10-31
;  PRIOR APPLICATION NUMBER: US 60/244,541
;  PRIOR FILING DATE: 2000-10-31
;  NUMBER OF SEQ ID NOS: 41
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12
;   LENGTH: 101
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: selected peptide
US-10-004-381-12

  Query Match             10.7%;  Score 6;  DB 14;  Length 101;
  Best Local Similarity   100.0%;  Pred. No. 54;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          15 QVGHDA 20
               ||||||
Db          31 QVGHDA 36


RESULT 7
US-10-029-386-34258
; Sequence 34258, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
;  TITLE OF INVENTION:  EXPRESSION ANALYSIS TWO
;  FILE REFERENCE: AEOMICA-X-2
;  CURRENT APPLICATION NUMBER: US/10/029,386
;  CURRENT FILING DATE:  2001-12-20
;  NUMBER OF SEQ ID NOS: 34288
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 34258
;   LENGTH: 157
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AL031709.10
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.7
;   OTHER INFORMATION: SWISSPROT HIT: P13677, EVALUE 2.60e-02
US-10-029-386-34258

  Query Match             10.7%;  Score 6;  DB 12;  Length 157;
  Best Local Similarity   100.0%;  Pred. No. 80;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          19 DANGEL 24
               ||||||
Db          90 DANGEL 95


RESULT 8
US-10-001-189-47
; Sequence 47, Application US/10001189
; Publication No. US20020173634A1
; GENERAL INFORMATION:
;  APPLICANT: FRASER JR., MALCOLM J.
;  APPLICANT:  LI, XU
;  APPLICANT:  BEAM, TERESA
;  TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR TRANSPOSITION USING
;  TITLE OF INVENTION:  MINIMAL SEGMENTS OF THE EUKARYOTIC TRANSFORMATION
;  TITLE OF INVENTION:  VECTOR PIGGYBAC
;  FILE REFERENCE: 835910-92098
;  CURRENT APPLICATION NUMBER: US/10/001,189
;  CURRENT FILING DATE:  2001-10-30
;  PRIOR APPLICATION NUMBER: 60/244,984
;  PRIOR FILING DATE: 2000-11-01
;  PRIOR APPLICATION NUMBER: 60/244,677
;  PRIOR FILING DATE: 2000-10-31
;  NUMBER OF SEQ ID NOS: 70
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 47
;   LENGTH: 193
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: pCRII-ITR
;   OTHER INFORMATION: amino acid sequence
US-10-001-189-47

  Query Match             10.7%;  Score 6;  DB 14;  Length 193;
  Best Local Similarity   100.0%;  Pred. No. 95;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          21 NGELFV 26
               ||||||
Db          73 NGELFV 78


RESULT 9
US-09-815-242-5521
; Sequence 5521, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
```

PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 5521
LENGTH: 202
TYPE: PRT
ORGANISM: Staphylococcus aureus
3-09-815-242-5521

Query Match 10.7%; Score 6; DB 9; Length 202;
Best Local Similarity 100.0%; Pred. No. 99;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 21 NGELFV 26
||||||
Db 158 NGELFV 163

RESULT 10
3-09-815-242-12448
Sequence 12448, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
TITLE OF INVENTION: Prokaryotes
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 12448
LENGTH: 202
TYPE: PRT
ORGANISM: Staphylococcus aureus
-09-815-242-12448

Query Match 10.7%; Score 6; DB 9; Length 202;
Best Local Similarity 100.0%; Pred. No. 99;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 21 NGELFV 26
||||||
Db 158 NGELFV 163

RESULT 11
US-09-966-521-30
; Sequence 30, Application US/09966521
; Publication No. US20030087821A1
; GENERAL INFORMATION:
; APPLICANT: TOMICH, Che-Shen
; APPLICANT: QUINN, Cheryl
; APPLICANT: ARVIDSON, Staffan
; APPLICANT: HARRIS, Douglas
; APPLICANT: MOTT, John
; TITLE OF INVENTION: ANTIMICROBIAL METHODS AND MATERIALS
; FILE REFERENCE: 6212.N2
; CURRENT APPLICATION NUMBER: US/09/966,521
; CURRENT FILING DATE: 2001-09-28
; NUMBER OF SEQ ID NOS: 145
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 30
; LENGTH: 202
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-966-521-30

Query Match 10.7%; Score 6; DB 11; Length 202;
Best Local Similarity 100.0%; Pred. No. 99;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 21 NGELFV 26
||||||
Db 158 NGELFV 163

RESULT 12
US-10-429-094-30
; Sequence 30, Application US/10429094
; Publication No. US20030180821A1
; GENERAL INFORMATION:
; APPLICANT: TOMICH, Che-Shen
; APPLICANT: QUINN, Cheryl
; APPLICANT: ARVIDSON, Staffan
; APPLICANT: HARRIS, Douglas
; APPLICANT: MOTT, John
; TITLE OF INVENTION: ANTIMICROBIAL METHODS AND MATERIALS
; FILE REFERENCE: 00774.US1 CN1
; CURRENT APPLICATION NUMBER: US/10/429,094
; CURRENT FILING DATE: 2003-05-02
; PRIOR APPLICATION NUMBER: US 09/966,521
; PRIOR FILING DATE: 2001-09-28
; NUMBER OF SEQ ID NOS: 145
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 30
; LENGTH: 202
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-10-429-094-30

Query Match 10.7%; Score 6; DB 12; Length 202;
Best Local Similarity 100.0%; Pred. No. 99;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 21 NGELFV 26
||||||
Db 158 NGELFV 163

RESULT 13
US-09-738-626-3592

; Sequence 3592, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 3592
; LENGTH: 308
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-3592

Query Match 10.7%; Score 6; DB 10; Length 308;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 51 CLRCGS 56
||||||
Db 210 CLRCGS 215

RESULT 14
US-09-726-643-72
; Sequence 72, Application US/09726643
; Patent No. US20020028449A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 26 Human secreted proteins
; FILE REFERENCE: PZ040P1
; CURRENT APPLICATION NUMBER: US/09/726,643
; CURRENT FILING DATE: 2000-12-01
; PRIOR APPLICATION NUMBER: PCT/US00/15187
; PRIOR FILING DATE: 2000-06-02
; PRIOR APPLICATION NUMBER: 60/137,725
; PRIOR FILING DATE: 1999-06-07
; NUMBER OF SEQ ID NOS: 190
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 72
; LENGTH: 322
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-726-643-72

Query Match 10.7%; Score 6; DB 9; Length 322;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 22 GELFVA 27
||||||
Db 117 GELFVA 122

RESULT 15
US-10-042-141-72
; Sequence 72, Application US/10042141
; Publication No. US20020183503A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 26 Human secreted proteins
; FILE REFERENCE: PZ040P1
; CURRENT APPLICATION NUMBER: US/10/042,141
; CURRENT FILING DATE: 2002-01-11
; PRIOR APPLICATION NUMBER: 09/726,643
; PRIOR FILING DATE: 2000-12-01
; PRIOR APPLICATION NUMBER: PCT/US00/15187
; PRIOR FILING DATE: 2000-06-02
; PRIOR APPLICATION NUMBER: 60/137,725
; PRIOR FILING DATE: 1999-06-07
; NUMBER OF SEQ ID NOS: 190
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 72
; LENGTH: 322
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-042-141-72

Query Match 10.7%; Score 6; DB 14; Length 322;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 22 GELFVA 27
||||||
Db 117 GELFVA 122

Search completed: October 29, 2003, 18:38:34
Job time : 4.97872 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

CM protein - protein search, using sw model

Run on: October 29, 2003, 18:33:14 ; Search time 1.62476 Seconds
(without alignments)
1458.314 Million cell updates/sec

Title: US-09-980-043-4
Perfect score: 56
Sequence: 1 MMESGAPICHTCGEQVGHDA.....KSCFEFEIKEGRKVCLRCGS 56

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 328717 seqs, 42310858 residues

Word size : 0

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	7	12.5	974	3	US-08-960-048-6	Sequence 6, Appli
2	7	12.5	974	4	US-09-838-586-6	Sequence 6, Appli
3	6	10.7	8	1	US-08-[illegible]-779-2	Sequence 7, Appli
4	6	10.7	8	1	US-08-798-946-2	Sequence 7, Appli
5	6	10.7	8	2	US-08-831-570-4	Sequence 4, Appli
6	6	10.7	8	2	US-08-831-575-14	Sequence 14, Appl
7	6	10.7	8	4	US-08-934-254-7	Sequence 7, Appli
8	6	10.7	213	3	US-08-943-173-3	Sequence 3, Appli
9	6	10.7	254	4	US-09-252-991A-30634	Sequence 30634, A
10	6	10.7	288	4	US-09-107-532A-7120	Sequence 7120, Ap
11	6	10.7	359	1	US-08-307-382-2	Sequence 2, Appli
12	6	10.7	359	1	US-08-366-779-2	Sequence 2, Appli
13	6	10.7	359	1	US-08-478-727-2	Sequence 2, Appli
14	6	10.7	359	1	US-08-473-508-2	Sequence 2, Appli
15	6	10.7	359	1	US-08-789-936-2	Sequence 2, Appli
16	6	10.7	359	2	US-08-833-610-6	Sequence 6, Appli
17	6	10.7	359	3	US-08-834-033A-16	Sequence 16, Appl
18	6	10.7	359	4	US-08-934-254-2	Sequence 2, Appli
19	6	10.7	390	4	US-09-328-352-7880	Sequence 7880, Ap
20	6	10.7	465	3	US-09-537-357-52	Sequence 52, Appl
21	6	10.7	466	4	US-09-252-991A-22137	Sequence 22137, A
22	6	10.7	475	4	US-09-252-991A-31373	Sequence 31373, A
23	6	10.7	476	4	US-09-252-991A-21886	Sequence 21886, A
24	6	10.7	605	2	US-08-752-307B-8	Sequence 8, Appli
25	6	10.7	605	4	US-09-707-802-8	Sequence 8, Appli
26	6	10.7	605	4	US-09-991-326-8	Sequence 8, Appli
27	6	10.7	614	4	US-09-328-352-4256	Sequence 4256, Ap
28	6	10.7	645	3	US-09-328-501-1	Sequence 1, Appli
29	6	10.7	646	4	US-09-777-710A-1	Sequence 1, Appli
30	6	10.7	650	4	US-09-134-001C-4444	Sequence 4444, Ap
31	6	10.7	658	4	US-09-252-991A-26319	Sequence 26319, A
32	6	10.7	670	3	US-09-328-501-15	Sequence 15, Appl
33	6	10.7	670	4	US-09-777-710A-15	Sequence 15, Appl
34	6	10.7	677	4	US-09-252-991A-20406	Sequence 20406, A
35	6	10.7	686	4	US-09-252-991A-29396	Sequence 29396, A
36	6	10.7	748	3	US-09-066-046-4	Sequence 4, Appli
37	6	10.7	879	4	US-09-107-532A-4679	Sequence 4679, Ap
38	6	10.7	1018	1	US-08-452-052-2	Sequence 2, Appli
39	6	10.7	1084	4	US-09-221-013A-8	Sequence 8, Appli
40	6	10.7	2318	3	US-09-091-219-24	Sequence 24, Appl
41	6	10.7	2318	4	US-09-660-541-24	Sequence 24, Appl
42	5	8.9	6	1	US-08-457-605-5	Sequence 5, Appli
43	5	8.9	6	1	US-08-457-605-7	Sequence 7, Appli
44	5	8.9	9	3	US-08-159-339A-919	Sequence 919, App
45	5	8.9	10	3	US-08-159-339A-936	Sequence 936, App

ALIGNMENTS

RESULT 1
US-08-960-048-6
; Sequence 6, Application US/08960048C
; Patent No. 6271443
; GENERAL INFORMATION:
; APPLICANT: Stalker, D. et al.
; TITLE OF INVENTION: Plant Cellulose Synthase and Promoter
; TITLE OF INVENTION: Sequences
; FILE REFERENCE: 15621/01/US
; CURRENT APPLICATION NUMBER: US/08/960,048C
; CURRENT FILING DATE: 1997-10-29
; PRIOR APPLICATION NUMBER: 60/029,987
; PRIOR FILING DATE: 1996-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 974
; TYPE: PRT
; ORGANISM: Gossypim hirsutum
US-08-960-048-6

Query Match 12.5%; Score 7; DB 3; Length 974;
Best Local Similarity 100.0%; Pred. No. 8.9;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 25 FVACHEC 31
|||||||
Db 25 FVACHEC 31

RESULT 2
US-09-838-586-6
; Sequence 6, Application US/09838586
; Patent No. 6576818
; GENERAL INFORMATION:
; APPLICANT: Stalker, D. et al.
; TITLE OF INVENTION: Plant Cellulose Synthase and Promoter
; TITLE OF INVENTION: Sequences
; FILE REFERENCE: 15621/02/US
; CURRENT APPLICATION NUMBER: US/09/838,586
; CURRENT FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/029,987
; PRIOR FILING DATE: 1996-10-29
; PRIOR APPLICATION NUMBER: 08/960,048
; PRIOR FILING DATE: 1997-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 974
; TYPE: PRT

ORGANISM: Gossypim hirsutum
S-09-838-586-6

Query Match 12.5%; Score 7; DB 4; Length 974;
Best Local Similarity 100.0%; Pred. No. 8.9;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 25 FVACHEC 31
||||||| |
Db 25 FVACHEC 31

ESULT 3
S-08-366-779-7
Sequence 7, Application US/08366779
Patent No. 5614393
GENERAL INFORMATION:
APPLICANT: Thomas, Terry L.
APPLICANT: Reddy, Avutu S.
APPLICANT: Nuccio, Michael
APPLICANT: Freyssinet, Georges L.
APPLICANT: Nunberg, Andrew N.
TITLE OF INVENTION: PRODUCTION OF GAMMA LINOLENIC ACID BY A
TITLE OF INVENTION: DELTA 6-DESATURASE
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Scully, Scott, Murphy & Presser
STREET: 400 Garden City Plaza
CITY: Garden City
STATE: New York
COUNTRY: United States
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/366,779
FILING DATE: 30-DEC-1994
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Presser, Leopold
REGISTRATION NUMBER: 19,827
REFERENCE/DOCKET NUMBER: 8383ZYXW
TELECOMMUNICATION INFORMATION:
TELEPHONE: (516) 742-4343
TELEFAX: (516) 742-4366
TELEX: 230 901 SANS UR
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 8 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
S-08-366-779-7

Query Match 10.7%; Score 6; DB 1; Length 8;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 16 VGHDAN 21
|||||| |
Db 2 VGHDAN 7

ESULT 4
S-08-789-936-7
Sequence 7, Application US/08789936
Patent No. 5789220
GENERAL INFORMATION:
APPLICANT: Thomas, Terry L.
APPLICANT: Reddy, Avutu S.
APPLICANT: Nuccio, Michael
APPLICANT: Freyssinet, Georges L.
APPLICANT: Nunberg, Andrew N.
TITLE OF INVENTION: PRODUCTION OF GAMMA LINOLENIC ACID BY A
TITLE OF INVENTION: DELTA 6-DESATURASE
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Scully, Scott, Murphy & Presser
STREET: 400 Garden City Plaza
CITY: Garden City
STATE: New York
COUNTRY: United States
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/789,936
FILING DATE: 28-JAN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/366,779
FILING DATE: 30-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: Presser, Leopold
REGISTRATION NUMBER: 19,827
REFERENCE/DOCKET NUMBER: 8383ZYXW
TELECOMMUNICATION INFORMATION:
TELEPHONE: (516) 742-4343
TELEFAX: (516) 742-4366
TELEX: 230 901 SANS UR
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 8 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE:
US-08-789-936-7

Query Match 10.7%; Score 6; DB 1; Length 8;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 16 VGHDAN 21
|||||| |
Db 2 VGHDAN 7

RESULT 5
US-08-831-570-4
; Sequence 4, Application US/08831570
; Patent No. 5959175
; GENERAL INFORMATION:
; APPLICANT: Thomas, Terry L.
; APPLICANT: Nunberg, Andrew N.
; APPLICANT: Beremand, Phillip D.
; TITLE OF INVENTION: A SUNFLOWER ALBUMIN 5' REGULATORY REGION
; TITLE OF INVENTION: FOR THE MODIFICATION OF PLANT SEED LIPID
; TITLE OF INVENTION: COMPOSITION
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Scully, Scott, Murphy & Presser
; STREET: 400 Garden City Plaza
; CITY: Garden City
; STATE: New York
; COUNTRY: USA
; ZIP: 11530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/831,570
;       FILING DATE:  09-APR-1997
;       CLASSIFICATION:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  DiGiglio, Frank S.
;       REGISTRATION NUMBER:  31,346
;       REFERENCE/DOCKET NUMBER:  10545
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (516) 742-4343
;       TELEFAX:  (516) 743-4366
;   INFORMATION FOR SEQ ID NO:  4:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  8 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-08-831-570-4

  Query Match             10.7%;  Score 6;  DB 2;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 2.5e+05;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          16 VGHDAN 21
               ||||||
Db           2 VGHDAN 7


RESULT 6
US-08-831-575-14
; Sequence 14, Application US/08831575
; Patent No. 5977436
;  GENERAL INFORMATION:
;     APPLICANT:  Thomas, Terry L.
;     APPLICANT:  Li, Zhongsen
;     TITLE OF INVENTION:  AN OLEOSIN 5'REGULATORY REGION FOR THE
;     TITLE OF INVENTION:  MODIFICATION OF PLANT SEED LIPID COMPOSITION
;     NUMBER OF SEQUENCES:  33
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  SCULLY, SCOTT, MURPHY & PRESSER
;       STREET:  400 Garden City Plaza
;       CITY:  Garden City
;       STATE:  New York
;       COUNTRY:  U.S.A
;       ZIP:  11530
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/831,575
;       FILING DATE:  09-APR-1997
;       CLASSIFICATION:  800
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  DiGiglio, Frank S.
;       REGISTRATION NUMBER:  31,346
;       REFERENCE/DOCKET NUMBER:  10208
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (516) 742-4343
;       TELEFAX:  (516) 742-4366
;   INFORMATION FOR SEQ ID NO:  14:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  8 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  peptide
US-08-831-575-14

  Query Match             10.7%;  Score 6;  DB 2;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 2.5e+05;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          16 VGHDAN 21
               ||||||
Db           2 VGHDAN 7


RESULT 7
US-08-934-254-7
; Sequence 7, Application US/08934254
; Patent No. 6355861
;  GENERAL INFORMATION:
;     APPLICANT:  Thomas, Terry L.
;     TITLE OF INVENTION:  PRODUCTION OF GAMMA LINOLENIC ACID BY A
;     TITLE OF INVENTION:  DELTA 6-DESATURASE
;     NUMBER OF SEQUENCES:  27
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Scully, Scott, Murphy & Presser
;       STREET:  400 Garden City Plaza
;       CITY:  Garden City
;       STATE:  New York
;       COUNTRY:  United States
;       ZIP:  11530
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/934,254
;       FILING DATE:
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Presser, Leopold
;       REGISTRATION NUMBER:  19,827
;       REFERENCE/DOCKET NUMBER:  8383ZYXWVU
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (516) 742-4343
;       TELEFAX:  (516) 742-4366
;       TELEX:  230 901 SANS UR
;   INFORMATION FOR SEQ ID NO:  7:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  8 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  Peptide
US-08-934-254-7

  Query Match             10.7%;  Score 6;  DB 4;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 2.5e+05;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          16 VGHDAN 21
               ||||||
Db           2 VGHDAN 7


RESULT 8
US-08-943-173-3
; Sequence 3, Application US/08943173
; Patent No. 6048538
;  GENERAL INFORMATION:
;     APPLICANT:  Wang, Chang Yi
;     APPLICANT:  Shen, Fan
;     APPLICANT:  Chen, Pei De
;     TITLE OF INVENTION:  PEPTIDES DERIVED
;     TITLE OF INVENTION:  FROM THE NON-STRUCTURAL PROTEINS OF
;     TITLE OF INVENTION:  FOOT AND MOUTH DISEASE VIRUS AS
;     TITLE OF INVENTION:  DIAGNOSTIC REAGENTS
;     NUMBER OF SEQUENCES:  16
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Maria C.H. Lin
```

STREET: 345 Park Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10154-0053
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/943,173
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Lin, Maria C.H.
REGISTRATION NUMBER: 29,323
REFERENCE/DOCKET NUMBER: 1151-4152
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-758-4800
TELEFAX: 212-751-6849
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 213 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
-08-943-173-3

Query Match 10.7%; Score 6; DB 3; Length 213;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

40 FEFEIK 45
||||||
70 FEFEIK 75

SULT 9
-09-252-991A-30634
Sequence 30634, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 30634
LENGTH: 254
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
-09-252-991A-30634

Query Match 10.7%; Score 6; DB 4; Length 254;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

26 VACHEC 31
||||||
54 VACHEC 59

SULT 10
-09-107-532A-7120
Sequence 7120, Application US/09107532A
Patent No. 6583275

; GENERAL INFORMATION:
; APPLICANT: Lynn A [illegible]otte-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/[illegible]
; FILING DATE: [illegible]
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: [illegible]
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-[illegible]
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 7120:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 288 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (B) LOCATION 1...288
; SEQUENCE DESCRIPTION: SEQ ID NO: 7120:
US-09-107-532A-7120

Query Match 10.7%; Score 6; DB 4; Length 288;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 ICHTCG 13
||||||
Db 53 ICHTCG 58

RESULT 11
US-08-307-382-2
; Sequence 2, Application US/08307382
; Patent No. 5552306
; GENERAL INFORMATION:
; APPLICANT: Thomas, Terry D.
; APPLICANT: Reddy, Avuth S.
; APPLICANT: Nuccio, Michael
; APPLICANT: Freyssinet, Georges L.
; TITLE OF INVENTION: PRODUCTION OF GAMMA LINOLENIC
; TITLE OF INVENTION: ACID BY A DELTA 6-DESATURASE
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Scully, Scott, Murphy & Presser
; STREET: 400 Garden City Plaza
; CITY: Garden City
; STATE: New York

```
;       COUNTRY:  United States
;       ZIP:  11530
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/307,382
;       FILING DATE:  14-SEP-1994
;       CLASSIFICATION:  800
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/817,919
;       FILING DATE:  08-JAN-1992
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  McNulty, William E.
;       REGISTRATION NUMBER:  22,606
;       REFERENCE/DOCKET NUMBER:  8383Z
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (516) 742-4343
;       TELEFAX:  (516) 742-4366
;       TELEX:  230 901 SANS UR
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  359 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-08-307-382-2

  Query Match             10.7%;  Score 6;  DB 1;  Length 359;
  Best Local Similarity   100.0%;  Pred. No. 50;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          16 VGHDAN 21
               ||||||
Db          86 VGHDAN 91


RESULT 12
US-08-366-779-2
; Sequence 2, Application US/08366779
; Patent No. 5614393
;  GENERAL INFORMATION:
;     APPLICANT:  Thomas, Terry L.
;     APPLICANT:  Reddy, Avutu S.
;     APPLICANT:  Nuccio, Michael
;     APPLICANT:  Freyssinet, Georges
;     APPLICANT:  Nunberg, Andrew N
;     TITLE OF INVENTION:  PRODUCTION OF GAMMA LINOLENIC ACID BY A
;     TITLE OF INVENTION:  DELTA 6-DESATURASE
;     NUMBER OF SEQUENCES:  25
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Scully, Scott, Murphy & Presser
;       STREET:  400 Garden City Plaza
;       CITY:  Garden City
;       STATE:  New York
;       COUNTRY:  United States
;       ZIP:  11530
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/366,779
;       FILING DATE:  30-DEC-1994
;       CLASSIFICATION:  800
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Presser, Leopold
;       REGISTRATION NUMBER:  19,827
;       REFERENCE/DOCKET NUMBER:  8383/YXW
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (516) 742-4343
;       TELEFAX:  (516) 742-4366
;       TELEX:  230 901 SANS UR
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  359 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-08-366-779-2

  Query Match             10.7%;  Score 6;  DB 1;  Length 359;
  Best Local Similarity   100.0%;  Pred. No. 50;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          16 VGHDAN 21
               ||||||
Db          86 VGHDAN 91


RESULT 13
US-08-478-727-2
; Sequence 2, Application US/08478727
; Patent No. 5663068
;  GENERAL INFORMATION:
;     APPLICANT:  Thomas, Terry L.
;     APPLICANT:  Reddy, Avutu S.
;     APPLICANT:  Nuccio, Michael
;     APPLICANT:  Freyssinet, Georges L.
;     TITLE OF INVENTION:  PRODUCTION OF GAMMA LINOLENIC
;     TITLE OF INVENTION:  ACID BY A DELTA 6-DESATURASE
;     NUMBER OF SEQUENCES:  3
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Scully, Scott, Murphy & Presser
;       STREET:  400 Garden City Plaza
;       CITY:  Garden City
;       STATE:  New York
;       COUNTRY:  United States
;       ZIP:  11530
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/478,727
;       FILING DATE:  07-JUN-1995
;       CLASSIFICATION:  800
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/307,382
;       FILING DATE:  14-SEP-1994
;       APPLICATION NUMBER:  US 07/817,919
;       FILING DATE:  08-JAN-1992
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  McNulty, William E.
;       REGISTRATION NUMBER:  22,606
;       REFERENCE/DOCKET NUMBER:  8383Z
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (516) 742-4343
;       TELEFAX:  (516) 742-4366
;       TELEX:  230 901 SANS UR
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  359 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-08-478-727-2

  Query Match             10.7%;  Score 6;  DB 1;  Length 359;
  Best Local Similarity   100.0%;  Pred. No. 50;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
        16 VGHDAN 21
           ||||||
        86 VGHDAN 91


SULT 14
-08-473-508-2
Sequence 2, Application US/08473508
Patent No. 5689050
 GENERAL INFORMATION:
   APPLICANT:  Thomas, Terry L.
   APPLICANT:  Reddy, Avutu S.
   APPLICANT:  Nuccio, Michael
   APPLICANT:  Freyssinet, Georges L.
   TITLE OF INVENTION:  PRODUCTION OF GAMMA LINOLENIC
   TITLE OF INVENTION:  ACID BY A DELTA 6-DESATURASE
   NUMBER OF SEQUENCES:  3
   CORRESPONDENCE ADDRESS:
     ADDRESSEE:  Scully, Scott, Murphy & Presser
     STREET:  400 Garden City Plaza
     CITY:  Garden City
     STATE:  New York
     COUNTRY:  United States
     ZIP:  11530
   COMPUTER READABLE FORM:
     MEDIUM TYPE:  Floppy disk
     COMPUTER:  IBM PC compatible
     OPERATING SYSTEM:  PC-DOS/MS-DOS
     SOFTWARE:  PatentIn Release #1.0, Version #1.25
   CURRENT APPLICATION DATA:
     APPLICATION NUMBER:  US/08/473,508
     FILING DATE:  07-JUN-1995
     CLASSIFICATION:  800
   PRIOR APPLICATION DATA:
     APPLICATION NUMBER:  US 08/307,382
     FILING DATE:  14-SEP-1994
     APPLICATION NUMBER:  US 07/817,919
     FILING DATE:  08-JAN-1992
   ATTORNEY/AGENT INFORMATION:
     NAME:  McNulty, William E.
     REGISTRATION NUMBER:  22,606
     REFERENCE/DOCKET NUMBER:  8383Z
   TELECOMMUNICATION INFORMATION:
     TELEPHONE:  (516) 742-4343
     TELEFAX:  (516) 742-4366
     TELEX:  230 901 SANS UR
 INFORMATION FOR SEQ ID NO:  2:
   SEQUENCE CHARACTERISTICS:
     LENGTH:  359 amino acids
     TYPE:  amino acid
     TOPOLOGY:  linear
   MOLECULE TYPE:  protein
-08-473-508-2

Query Match             10.7%;  Score 6;  DB 1;  Length 359;
Best Local Similarity   100.0%;  Pred. No. 50;
Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

        16 VGHDAN 21
           ||||||
        86 VGHDAN 91


SULT 15
-08-789-936-2
Sequence 2, Application US/08789936
Patent No. 5789220
 GENERAL INFORMATION:
   APPLICANT:  Thomas, Terry L.
   APPLICANT:  Reddy, Avutu S.
   APPLICANT:  Nuccio, Michael
   APPLICANT:  Freyssinet, Georges L.
   APPLICANT:  Nunberg, Andrew N.
;    TITLE OF INVENTION:  PRODUCTION OF GAMMA LINOLENIC ACID BY A
;    TITLE OF INVENTION:  DELTA 6-DESATURASE
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Scully, Scott, Murphy & Presser
;      STREET:  400 Garden City Plaza
;      CITY:  Garden City
;      STATE:  New York
;      COUNTRY:  United States
;      ZIP:  11530
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/789,936
;      FILING DATE:  28-JAN-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/307,382
;      FILING DATE:  30-DEC-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Presser, Leopold
;      REGISTRATION NUMBER:  19,827
;      REFERENCE/DOCKET NUMBER:  8383ZYXW
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (516) 742-4343
;      TELEFAX:  (516) 742-4366
;      TELEX:  230 901 SANS UR
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  359 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-789-936-2

  Query Match             10.7%;  Score 6;  DB 1;  Length 359;
  Best Local Similarity   100.0%;  Pred. No. 50;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          16 VGHDAN 21
               ||||||
Db          86 VGHDAN 91


Search completed: October 29, 2003, 18:37:32
Job time : 2.62476 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 29, 2003, 18:29:38 ; Search time 4.33269 Seconds
(without alignments)
2051.541 Million cell updates/sec

Title: US-09-980-043-4
Perfect score: 56
Sequence: 1 MMESGAPICHTCGEQVGHDA.....KSCFEFEIKEGRKVCLRCGS 56

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1107863 seqs, 158726573 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	56	100.0	56	22	AAB37894	Populus tremuloide
2	14	25.0	946	22	AAB37893	Populus tremuloide
3	8	14.3	245	21	AAB16308	Pinus radiata cell
4	8	14.3	958	23	ABB92981	Herbicidally activ
5	7	12.5	44	19	AAW98285	H. pylori GHPO 131
6	7	12.5	184	21	AAB16320	Eucalyptus grandis
7	7	12.5	216	23	AAU83219	Novel secreted pro
8	7	12.5	251	21	AAB16274	Eucalyptus grandis
9	7	12.5	312	21	AAB18153	Plasmodium falcipa
10	7	12.5	603	21	AAB16307	Pinus radiata cell
11	7	12.5	728	22	ABB65430	Drosophila melanog
12	7	12.5	955	21	AAB16338	Eucalyptus grandis
13	7	12.5	974	19	AAW73308	Cellulose synthase
14	7	12.5	974	19	AAW60039	Cotton cellulose s
15	7	12.5	2825	21	AAG30671	Arabidopsis thalia
16	7	12.5	2837	21	AAG30670	Arabidopsis thalia
17	7	12.5	2895	21	AAG30669	Arabidopsis thalia
18	6	10.7	8	24	ABG73097	Membrane-bound des
19	6	10.7	40	23	AAE22588	BS203 consensus se
20	6	10.7	70	22	ABG26387	Novel human diagno
21	6	10.7	82	24	ABP75649	Human secretory po
22	6	10.7	101	23	ABG67066	Streptavidin-bindi
23	6	10.7	101	24	ABU57588	Streptavidin-bindi
24	6	10.7	116	22	AAM84104	Human immune/haema
25	6	10.7	124	22	ABG01207	Novel human diagno
26	6	10.7	147	22	ABG11460	Novel human diagno
27	6	10.7	159	22	AAM25942	Human protein sequ
28	6	10.7	173	22	ABG04008	Novel human diagno
29	6	10.7	193	24	ABU70357	Transformation vec
30	6	10.7	193	24	ABU70358	Transformation vec
31	6	10.7	202	22	AAU34025	Staphylococcus aur
32	6	10.7	202	22	AAU36855	Staphylococcus aur
33	6	10.7	213	21	AAY85078	FMDV non-structura
34	6	10.7	229	23	ABB49336	Listeria monocytog
35	6	10.7	250	23	ABB48663	Listeria monocytog
36	6	10.7	256	23	ABB55350	Lactococcus lactis
37	6	10.7	256	23	ABB47643	Listeria monocytog
38	6	10.7	262	22	AAU25441	Human mddt protein
39	6	10.7	269	22	ABB66307	Drosophila melanog
40	6	10.7	277	9	AAP80437	Gentamicin-3-N-ace
41	6	10.7	298	21	AAG10292	Arabidopsis thalia
42	6	10.7	298	21	AAG22235	Arabidopsis thalia
43	6	10.7	298	21	AAG45387	Arabidopsis thalia
44	6	10.7	308	22	AAG89838	C glutamicum prote
45	6	10.7	322	22	AAB65019	Human secreted pro

ALIGNMENTS

RESULT 1
AAB37894
ID AAB37894 standard; Protein; 56 AA.
XX
AC AAB37894;
XX
DT 07-MAR-2001 (first entry)
XX
DE Populus tremuloides CelA N-terminal sequence.
XX
KW Aspen; CelA; cellulose synthase; arabidopsis; wood;
KW mechanical stress response element; MSRE; cellulose biosynthesis.
XX
OS Populus tremuloides.
XX
PN WO200071670-A2.
XX
PD 30-NOV-2000.
XX
PF 18-MAY-2000; 2000WO-US13637.
XX
PR 21-MAY-1999; 99US-0135280.
XX
PA (UNMT) UNIV MICHIGAN TECHNOLOGICAL.
XX
PI Chiang VL, Wu L, Chandrashekhar JP, Carraway DT;
XX
DR WPI; 2001-032025/04.
DR N-PSDB; AAC65449.
XX
PT Novel polynucleotide encoding cellulose synthase useful for enhancing
PT cellulose biosynthesis, modifying lignin biosynthesis, and for

accelerating growth of plants and for upregulating cellulose synthase levels -

Claim 1; Fig 4; 65pp; English.

The present sequence is given in a specification relating to polynucleotides encoding cellulose synthase enzymes from Populus tremuloides and Arabidopsis. The cellulose synthases are designated PtCelA and AraxCelA respectively. The polynucleotides may be introduced into plants where they are useful for altering the growth of a plant. Vectors comprising a cellulose synthase promoter, or a fragment of the promoter containing one or more mechanical stress response elements (MSRE), operably linked to a cellulose synthase gene are useful for causing stress-induced gene expression in a plant cell. A cDNA encoding a protein that binds to a positive MSRE may be delivered into the plant to increase expression of cellulose synthase. A cDNA in an antisense orientation may be delivered to reduce expression of cellulose synthase. The polynucleotides enable the development of transgenic tree varieties having increased cellulose content, decreased lignin content and therefore improved wood fibre characteristics. The cellulose synthase promoters are useful for synthesising greater quantities of high-crystalline cellulose in plant and preferably in trees, which permits an increase in proportion of cellulose in transgenic plants, greater strength of juvenile wood and fiber and acceleration of overall growth rate.
Note: This polypeptide sequence is given in the specification as SEQ ID NO: 4. SEQ ID NO: 4 is referred to in claim 1 but it is described as a nucleotide sequence.

Sequence 56 AA;

```
Query Match             100.0%;  Score 56;  DB 22;  Length 56;
Best Local Similarity   100.0%;  Pred. No. 5.3e-54;
Matches   56;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

          1 MMESGAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEIKEGRKVCLRCGS 56
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
          1 MMESGAPICHTCGEQVGHDANGELFVACHECSYPMCKSCFEFEIKEGRKVCLRCGS 56
```

RESULT 2
AAB37893
AAB37893 standard; Protein; 946 AA.

AAB37893;

07-MAR-2001 (first entry)

Populus tremuloides cellulose synthase CelA.

Aspen; CelA; cellulose synthase; arabidopsis; wood; mechanical stress response element; MSRE; cellulose biosynthesis.

Populus tremuloides.

WO200071670-A2.

30-NOV-2000.

18-MAY-2000; 2000WO-US13637.

21-MAY-1999; 99US-0135280.

(UNMT) UNIV MICHIGAN TECHNOLOGICAL.

Chiang VL, Wu L, Chandrashekhar JP, Carraway DT;

WPI; 2001-032025/04.
N-PSDB; AAC65448.

Novel polynucleotide encoding cellulose synthase useful for enhancing cellulose biosynthesis, modifying lignin biosynthesis, and for accelerating growth of plants and for upregulating cellulose synthase levels -

Claim 8; Fig 1; 65pp; English.

The present sequence is given in a specification relating to polynucleotides encoding cellulose synthase enzymes from Populus tremuloides and Arabidopsis. The cellulose synthases are designated PtCelA and AraxCelA respectively. The polynucleotides may be introduced into plants where they are useful for altering the growth of a plant. Vectors comprising a cellulose synthase promoter, or a fragment of the promoter containing one or more mechanical stress response elements (MSRE), operably linked to a cellulose synthase gene are useful for causing stress-induced gene expression in a plant cell. A cDNA encoding a protein that binds to a positive MSRE may be delivered into the plant to increase expression of cellulose synthase. A cDNA in an antisense orientation may be delivered to reduce expression of cellulose synthase. The polynucleotides enable the development of transgenic tree varieties having increased cellulose content, decreased lignin content and therefore improved wood fibre characteristics. The cellulose synthase promoters are useful for synthesising greater quantities of high-crystalline cellulose in plant and preferably in trees, which permits an increase in proportion of cellulose in transgenic plants, greater strength of juvenile wood and fiber and acceleration of overall growth rate.

Sequence 946 AA;

```
Query Match             25.0%;  Score 14;  DB 22;  Length 946;
Best Local Similarity   100.0%;  Pred. No. 9.6e-07;
Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MMESGAPICHTCGE 14
              ||||||||||||||
Db          1 MMESGAPICHTCGE 14
```

RESULT 3
AAB16308
AAB16308 standard; Protein; 245 AA.

AAB16308;

31-OCT-2000 (first entry)

Pinus radiata cellulose synthase protein sequence SEQ ID NO:94.

Eucalyptus grandis; pinus radiata; Monterey pine; modification; plant cell wall; polysaccharide; polysaccharide biosynthetic pathway; transgenic plant.

Pinus radiata.

WO200022092-A2.

20-APR-2000.

08-OCT-1999; 99WO-NZ00169.

13-OCT-1998; 98US-0170862.
11-AUG-1999; 99US-0148426.

(GENE-) GENESIS RES & DEV CORP LTD.
(FLET-) FLETCHER CHALLENGE FORESTS LTD.

Bloksberg LN;

WPI; 2000-339328/29.
N-PSDB; AAA67115.

New genes encoding proteins involved in a plant polysaccharide biosynthetic pathway, useful for modulating or altering the polysaccharide content, composition or structure of the plant -

PS Claim 17; Page 84-85; 301pp; English.
XX
CC The present invention describes isolated polynucleotides (PN) comprising
CC a sequence selected from one of 835 nucleotide sequences given in
CC AAA67073 to AAA67907, their (reverse) complements, sequences producing
CC an Expectation (E) value of 0.01 or less compared to the 835 sequences,
CC sequences at least 50% identical to them, 200, 100, 40 or 20-mers of the
CC 835 sequences or sequences that are degenerately equivalent or allelic
CC to the 835 sequences. The polynucleotides are used to modify the
CC activity of a polypeptide involved in a polysaccharide biosynthetic
CC pathway in the plant. They are especially used to modulate or alter the
CC polysaccharide content, composition or structure of the plant. AAB16268
CC to AAB16340 are proteins encoded by some of the polynucleotide sequence
CC given in the present invention.
XX
SQ Sequence 245 AA;

Query Match 14.3%; Score 8; DB 21; Length 245;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 21 NGELFVAC 28
||||||||
Db 65 NGELFVAC 72

RESULT 4
ABB92981
ID ABB92981 standard; Protein; 958 AA.
XX
AC ABB92981;
XX
DT 31-MAY 2002 (first entry)
XX
DE Herbicidally active polypeptide SEQ ID NO 2192.
XX
KW Herbicidal; plant; agriculture; herbicide.
XX
CS Arabidopsis thaliana.
XX
PN WO200210210-A2.
XX
PD 07-FEB-2002.
XX
PF 28-AUG-2001; 2001WO-EP09892.
XX
PR 28-AUG-2001; 2001WO-EP09892.
XX
PA (FARB) BAYER AG.
XX
PI Tietjen K, Weidler M;
XX
DR WPI; 2002-269010/31.
XX
PT Identifying plant target proteins for herbicidally active compounds,
PT comprising aligning and comparing nucleic acid or amino acid sequences
PT from plant with nucleic acid or amino acid sequences from non-plant
PT organisms -
XX
PS Claim 5; SEQ ID NO 2192; 261pp + Sequence Listing; English.
XX
CC The invention relates to identifying target proteins
CC (ABB90790-ABB94016) for herbicidally active compounds, comprising
CC aligning and comparing nucleic acid or amino acid sequences from plant
CC with nucleic acid or amino acid sequences from non plant organisms using
CC suitable search parameters, where plant sequences having an E-value
CC greater by a factor of 3 than the E-value of most similar non-plant
CC sequences are selected. The polypeptides or nucleic acids encoding them
CC are useful for identifying modulators. The identified modulators are
CC useful as herbicides.
XX
SQ Sequence 958 AA;

Query Match 14.3%; Score 8; DB 23; Length 958;
Best Local Similarity 100.0%; Pred. No. 3.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 25 FVACHECS 32
||||||||
Db 25 FVACHECS 32

RESULT 5
AAW98285
ID AAW98285 standard; Protein; 44 AA.
XX
AC AAW98285;
XX
DT 31-MAR-1999 (first entry)
XX
DE H. pylori GHPO 1316 protein.
XX
KW GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;
KW peptic ulcer disease.
XX
OS Helicobacter pylori.
XX
PN WO9843478-A1.
XX
PD 08-OCT-1998.
XX
PF 01-APR-1998; 98WO-US06371.
XX
PR 29-JUL-1997; 97US-0902615.
PR 01-APR-1997; 97US-0833457.
PR 24-JUN-1997; 97US-0881227.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
XX
PI Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;
XX
DR WPI; 1998-542293/46.
DR N-PSDB; AAX14004.
XX
PT New isolated Helicobacter polynucleotides - used to develop products
PT for the diagnosis, prevention and treatment of Helicobacter
PT infections and gastrointestinal diseases
XX
PS Claim 8; Page 338; 2054pp; English.
XX
CC This sequence represents a Helicobacter pylori GHPO protein of the
CC invention. The polypeptides can be used for preventing or treating
CC Helicobacter infections, and gastroduodenal diseases associated with
CC these infections, including acute, chronic, and atrophic gastritis, and
CC peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be
CC used for the production of antibodies. The products can also be used for
CC detection and diagnosis.
XX
SQ Sequence 44 AA;

Query Match 12.5%; Score 7; DB 19; Length 44;
Best Local Similarity 100.0%; Pred. No. 3.1;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 24 LFVACHE 30
|||||||
Db 13 LFVACHE 19

RESULT 6
AAB16320
ID AAB16320 standard; Protein; 184 AA.
XX
AC AAB16320;
XX

31-OCT-2000 (first entry)

Eucalyptus grandis cellulose synthase protein sequence SEQ ID NO:108.

Eucalyptus grandis; pinus radiata; Monterey pine; modification;
plant cell wall; polysaccharide; polysaccharide biosynthetic pathway;
transgenic plant.

Eucalyptus grandis.

WO200022092-A2.

20-APR-2000.

08-OCT-1999; 99WO-NZ00169.

13-OCT-1998; 98US-0170862.
11-AUG-1999; 99US-0148426.

(GENE-) GENESIS RES & DEV CORP LTD.
(FLET-) FLETCHER CHALLENGE FORESTS LTD.

Bloksberg LN;

WPI; 2000-339328/29.
N-PSDB; AAA67127.

New genes encoding proteins involved in a plant polysaccharide
biosynthetic pathway, useful for modulating or altering the
polysaccharide content, composition or structure of the plant -

Claim 17; Page 92-93; 301pp; English.

The present invention describes isolated polynucleotides (PN) comprising
a sequence selected from one of 835 nucleotide sequences given in
AAA67073 to AAA67907, their (reverse) complements, sequences producing
an Expectation (E) value of 0.01 or less compared to the 835 sequences,
sequences at least 50% identical to them, 200, 100, 40 or 20-mers of the
835 sequences or sequences that are degenerately equivalent or allelic
to the 835 sequences. The polynucleotides are used to modify the
activity of a polypeptide involved in a polysaccharide biosynthetic
pathway in the plant. They are especially used to modulate or alter the
polysaccharide content, composition or structure of the plant. AAB16268
to AAB16340 are proteins encoded by some of the polynucleotide sequence
given in the present invention.

Sequence 184 AA;

Query Match 12.5%; Score 7; DB 21; Length 184;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

43 EIKEGRK 49
43 EIKEGRK 49

RESULT 7
AAU83219
ID AAU83219 standard; Protein; 216 AA.
XX
AC AAU83219;
XX
DT 08-MAY-2002 (first entry)
XX
DE Novel secreted protein Z910197G6P.
XX
KW Protein secretion; mammalian secreted polypeptide; MSP.
XX
OS Homo sapiens.
XX
PN WO200202621-A2.
XX
PD 10-JAN-2002.
XX
PF 28-JUN-2001; 2001WO-US20848.
XX
PR 30-JUN-2000; 2000US-215446P.
XX
PA (ZYMO) ZYMOGENETICS INC.
XX
PI Sheppard PO, Presnell SR;
XX
DR WPI; 2002-147999/19.
DR N-PSDB; ABK33194.
XX
PT Novel isolated mammalian secreted polypeptide useful in therapeutic and
PT diagnostic methods, to direct secretion of other proteins of interest
PT from host cell, as educational tools, and as laboratory practicum kits
PT -
XX
PS Claim 12; Page 452-454; 670pp; English.
XX
CC The invention describes an isolated mammalian secreted polypeptide (MSP)
CC (I). (I) is useful to direct the secretion of other proteins of interest
CC from a host cell, to mediate secretion of proteins, to degenerate
CC sequences comprising all nucleotide sequences encoding a particular
CC polypeptide, to screen for cell metabolism effecting receptors, for
CC identifying new target receptors and drug design, for identifying, for
CC protein purification, for determining the weight of expressed MSP
CC polypeptides as a ratio to total protein expressed, for identifying
CC peptide cleavage sites, for coupling amino and carboxy terminal tags, for
CC amino acid sequence analysis, for monitoring biological activities of the
CC protein in vitro and in vivo, and to teach analytical skills and as
CC reagents for the study of cells, receptors, and other binding molecules.
CC The polynucleotide is useful for radiation hybrid mapping, and somatic
CC cell genetic technique developed for constructing high-resolution,
CC contiguous maps of mammalian chromosomes. Reagents disclosed in the
CC invention may be used to detect metabolic abnormalities characterised by
CC over or under production of the protein. This is the amino acid sequence
CC of a mammalian secreted polypeptide, described in the method of the
CC invention.
XX
SQ Sequence 216 AA;

Query Match 12.5%; Score 7; DB 23; Length 216;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 ESGAPIC 9
Db 151 ESGAPIC 157

RESULT 8
AAB16274
ID AAB16274 standard; Protein; 251 AA.
XX
AC AAB16274;
XX
DT 31-OCT-2000 (first entry)
XX
DE Eucalyptus grandis cellulose synthase protein sequence SEQ ID NO:36.
XX
KW Eucalyptus grandis; pinus radiata; Monterey pine; modification;
KW plant cell wall; polysaccharide; polysaccharide biosynthetic pathway;
KW transgenic plant.
XX
OS Eucalyptus grandis.
XX
PN WO200022092-A2.
XX
PD 20-APR-2000.
XX
PF 08-OCT-1999; 99WO-NZ00169.
XX

PR 13-OCT-1998; 98US-0170862.
PR 11-AUG-1999; 99US-0148426.
XX
PA (GENE-) GENESIS RES & DEV CORP LTD.
PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI Bloksberg LN;
XX
DR WPI; 2000-339328/29.
DR N-PSDB; AAA67079.
XX
PT New genes encoding proteins involved in a plant polysaccharide
PT biosynthetic pathway, useful for modulating or altering the
PT polysaccharide content, composition or structure of the plant -
XX
PS Claim 17; Page 52-53; 301pp; English.
XX
CC The present invention describes isolated polynucleotides (PN) comprising
CC a sequence selected from one of 835 nucleotide sequences given in
CC AAA67073 to AAA67907, their (reverse) complements, sequences producing
CC an Expectation (E) value of 0.01 or less compared to the 835 sequences,
CC sequences at least 50% identical to them, 200, 100, 40 or 20-mers of the
CC 835 sequences or sequences that are degenerately equivalent or allelic
CC to the 835 sequences. The polynucleotides are used to modify the
CC activity of a polypeptide involved in a polysaccharide biosynthetic
CC pathway in the plant. They are especially used to modulate or alter the
CC polysaccharide content, composition or structure of the plant. AAB16268
CC to AAB16340 are proteins encoded by some of the polynucleotide sequence
CC given in the present invention.
XX
SQ Sequence 251 AA;

Query Match 12.5%; Score 7; DB 21; Length 251;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 43 EIKEGRK 49
|||||||
Db 43 EIKEGRK 49

RESULT 9
AAB18153
ID AAB18153 standard; Protein; 312 AA.
XX
AC AAB18153;
XX
DT 07 NOV 2000 (first entry)
XX
DE Plasmodium falciparum chromosome 2 related protein SEQ ID NO:10.
XX
KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
KW antimalarial; malaria; protozoacide; infection; insecticide.
XX
OS Plasmodium falciparum.
XX
PN WO200025728-A2.
XX
PD 11 MAY-2000.
XX
PF 05-NOV-1999; 99WO-US26796
XX
PR 05-NOV-1998; 98US-0107131.
XX
PA (HOFF/) HOFFMAN S.
PA (CARU/) CARUCCI D.
PA (GARD/) GARDNER M.
PA (VENT/) VENTER J C.
XX
PI Hoffman S, Carucci D, Gardner M, Venter JC;
XX
DR WPI; 2000-365347/31.
XX
PT Proteins encoded by chromosome 2 of the human malarial parasite,
PT Plasmodium falciparum, useful as antimalarial vaccines and in the
PT diagnosis of P.falciparum infection -
XX
PS Disclosure; Page 41-42; 577pp; English.
XX
CC The present invention describes proteins and their fragments (I) encoded
CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
CC Also described are: (1) nucleotide sequences (II) encoding (I); and (2)
CC vaccines against P. falciparum infection comprising (I) or (II).
CC (I) and (II) are useful for the development of vaccines against
CC P. falciparum infection. (I) and polyclonal antisera or a monoclonal
CC antibody raised to immunogens comprising the sequences of (I), are
CC useful in the detection of infection with P. falciparum. Furthermore,
CC (I) (especially when they are rifins or secreted or membrane proteins)
CC can aid the identification of drugs to treat or prevent P. falciparum
CC infection, or they can be used to identify drug resistance in
CC P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
CC subsequent identification of proteins encoded by it will help to expand
CC our understanding of parasite biology, a process hampered by the
CC complexity of the parasitic lifecycle, and provide new targets for
CC vaccine and drug development. Parasite resistance to drugs and mosquito
CC resistance to insecticides have led to a resurgence of malaria in many
CC parts of the world, and there is a pressing need for vaccines and new
CC drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352 represent nucleotide
CC and protein sequences given in the present invention, but which are not
CC specifically mentioned within the specification.
XX
SQ Sequence 312 AA;

Query Match 12.5%; Score 7; DB 21; Length 312;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 49 KVCLRCG 55
|||||||
Db 113 KVCLRCG 119

RESULT 10
AAB16307
ID AAB16307 standard; Protein; 603 AA.
XX
AC AAB16307;
XX
DT 31-OCT-2000 (first entry)
XX
DE Pinus radiata cellulose synthase protein sequence SEQ ID NO:93.
XX
KW Eucalyptus grandis; pinus radiata; Monterey pine; modification;
KW plant cell wall; polysaccharide; polysaccharide biosynthetic pathway;
KW transgenic plant.
XX
OS Pinus radiata.
XX
PN WO200022092-A2.
XX
PD 20-APR-2000.
XX
PF 08-OCT-1999; 99WO-NZ00169.
XX
PR 13-OCT-1998; 98US-0170862.
PR 11-AUG-1999; 99US-0148426.
XX
PA (GENE-) GENESIS RES & DEV CORP LTD.
PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI Bloksberg LN;
XX
DR WPI; 2000-339328/29.
DR N-PSDB; AAA67114.
XX
PT New genes encoding proteins involved in a plant polysaccharide

biosynthetic pathway, useful for modulating or altering the polysaccharide content, composition or structure of the plant

Claim 17; Page 83-84; 301pp; English.

The present invention describes isolated polynucleotides (PN) comprising a sequence selected from one of 835 nucleotide sequences given in AAA67073 to AAA67907, their (reverse) complements, sequences producing an Expectation (E) value of 0.01 or less compared to the 835 sequences, sequences at least 50% identical to them, 200, 100, 40 or 20 mers of the 835 sequences or sequences that are degenerately equivalent or allelic to the 835 sequences. The polynucleotides are used to modify the activity of a polypeptide involved in a polysaccharide biosynthetic pathway in the plant. They are especially used to modulate or alter the polysaccharide content, composition or structure of the plant. AAB16268 to AAB16340 are proteins encoded by some of the polynucleotide sequence given in the present invention.

Sequence 603 AA;

```
Query Match             12.5%;  Score 7;  DB 21;  Length 603;
Best Local Similarity   100.0%;  Pred. No. 32;
Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

        22 GELFVAC 28
           |||||||
        63 GELFVAC 69
```

SULT 11
B65430
ABB65430 standard; Protein; 728 AA.

ABB65430;

26-MAR-2002 (first entry)

Drosophila melanogaster polypeptide SEQ ID NO 23082.

Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical.

Drosophila melanogaster.

WO200171042-A2.

27-SEP-2001.

23-MAR-2001; 2001WO-US09231.

23-MAR-2000; 2000US-191637P.
11-JUL-2000; 2000US-0614150.

(PEKE) PE CORP NY.

Venter JC, Adams M, Li PWD, Myers EW;

WPI; 2001-656860/75.
N-PSDB; ABL09533.

New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -

Disclosure; SEQ ID NO 23082; 21pp + Sequence Listing; English.

The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-ABB72072).
The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 728 AA;

```
  Query Match             12.5%;  Score 7;  DB 22;  Length 728;
  Best Local Similarity   100.0%;  Pred. No. 38;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          8 ICHTCGE 14
              |||||||
Db        515 ICHTCGE 521
```

RESULT 12
AAB16338
ID AAB16338 standard; Protein; 955 AA.
AC AAB16338;
DT 31-OCT-2000 (first entry)
DE Eucalyptus grandis cellulose synthase protein sequence SEQ ID NO:146.
KW Eucalyptus grandis; pinus radiata; Monterey pine; modification; plant cell wall; polysaccharide; polysaccharide biosynthetic pathway; transgenic plant.
OS Eucalyptus grandis.
PN WO200022092-A2
PD 20-APR-2000.
PF 08-OCT-1999; 99WO NZ00169.
PR 13-OCT-1998; 98US-0170862.
PR 11-AUG-1999; 99US-0148426.
PA (GENE-) GENESIS RES & DEV CORP LTD.
PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
PI Bloksberg LN;
DR WPI; 2000-339328/29.
DR N-PSDB; AAA67145.
PT New genes encoding proteins involved in a plant polysaccharide biosynthetic pathway, useful for modulating or altering the polysaccharide content, composition or structure of the plant -
PS Claim 17; Page 111-113; 301pp; English.
CC The present invention describes isolated polynucleotides (PN) comprising a sequence selected from one of 835 nucleotide sequences given in AAA67073 to AAA67907, their (reverse) complements, sequences producing an Expectation (E) value of 0.01 or less compared to the 835 sequences, sequences at least 50% identical to them, 200, 100, 40 or 20-mers of the 835 sequences or sequences that are degenerately equivalent or allelic to the 835 sequences. The polynucleotides are used to modify the activity of a polypeptide involved in a polysaccharide biosynthetic pathway in the plant. They are especially used to modulate or alter the polysaccharide content, composition or structure of the plant. AAB16268 to AAB16340 are proteins encoded by some of the polynucleotide sequence given in the present invention.
SQ Sequence 955 AA;

```
  Query Match             12.5%;  Score 7;  DB 21;  Length 955;
  Best Local Similarity   100.0%;  Pred. No. 48;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy          43 EIKEGRK 49
               |||||||
Db          43 EIKEGRK 49


RESULT 13
AAW73308
ID   AAW73308 standard; Protein; 974 AA.
XX
AC   AAW73308;
XX
DT   08-FEB-1999  (first entry)
XX
DE   Cellulose synthase PcsA1.
XX
KW   Cellulose synthase; cotton.
XX
OS   Gossypium hirsutum.
XX
PN   EP875575-A2.
XX
PD   04-NOV-1998.
XX
PF   31-MAR-1998;   98EP-0302489.
XX
PR   01-APR-1997;   97JP-0083133.
XX
PA   (NISN ) NISSHINBO IND INC.
XX
PI   Aotsuka S,  Hasegawa O,  Hayashi T,  Ihara Y;
XX
DR   WPI; 1998-559440/48.
DR   N-PSDB; AAV08372.
XX
PT   New DNA encoding cotton cellulose synthase proteins - useful for
PT   controlling cellulose synthesis in prokaryotic or eukaryotic cells
XX
PS   Claim 1; Page 14-18; 51pp; English.
XX
CC   This sequence is a cellulose synthase of the invention. The DNA can
CC   be used for controlling cellulose synthesis in a cell by introducing the
CC   DNA into the cell, expressing RNA with a nucleotide sequence homologous
CC   to the DNA or a nucleotide sequence complementary to the DNA. The cells
CC   are prokaryotic cells i.e. acetobacterium and/or eukaryotic cells such as
CC   yeasts i.e. Saccharomyces spp., plant cells (cotton plants) and cultured
CC   cells of mammals etc.
XX
SQ   Sequence   974 AA;

  Query Match             12.5%;  Score 7;  DB 19;  Length 974;
  Best Local Similarity   100.0%;  Pred. No. 49;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          25 FVACHEC 31
               |||||||
Db          25 FVACHEC 31


RESULT 14
AAW60039
ID   AAW60039 standard; Protein; 974 AA.
XX
AC   AAW60039;
XX
DT   28-SEP-1998  (first entry)
XX
DE   Cotton cellulose synthase CelA1.
XX
KW   Cellulose synthase; CelA1; cotton; forest tree; plant fibre;
KW   wood; transgenic plant; ss.
XX
OS   Gossypium hirsutum Acala SJ-2.
XX
FH   Key             Location/Qualifiers
FT   Region          175..295
FT                   /label= H-1
FT   Region          280..295
FT                   /label= U-1
FT   Region          296..420
FT                   /label= P-CR
FT   Region          421..539
FT                   /label= H-2
FT   Region          450..467
FT                   /label= U-2
FT   Region          540..658
FT                   /label= HVR
FT   Region          659..880
FT                   /label= H-3
FT   Region          667..688
FT                   /label= U-3
FT   Region          706..722
FT                   /label= U-4
FT   Modified-site   81
FT                   /note= "N-glycosylated"
FT   Modified-site   466
FT                   /note= "N-glycosylated"
FT   Modified-site   703
FT                   /note= "N-glycosylated"
FT   Modified-site   846
FT                   /note= "N-glycosylated"
FT   Misc-difference 293
FT                   /note= "critical Asp residue"
FT   Misc-difference 461
FT                   /note= "critical Asp residue"
FT   Misc-difference 672
FT                   /note= "critical Asp residue"
FT   Peptide         710..714
FT                   /note= "QXXRW motif"
XX
PN   WO9818949-A2.
XX
PD   07-MAY-1998.
XX
PF   29-OCT-1997;   97WO-US19529.
XX
PR   29-OCT-1996;   96US-0029987.
XX
PA   (CALJ ) CALGENE INC.
XX
PI   Delmer DP,  Pear JR,  Stalker DM;
XX
DR   WPI; 1998-272238/24.
DR   N-PSDB; AAV34432.
XX
PT   New isolated plant cellulose synthesis enzyme DNA - is used to
PT   develop products for producing modified plant fibre products,
PT   particularly for modifying cotton fibre and wood quality
XX
PS   Example 5; Fig 3A-C; 49pp; English.
XX
CC   This full-length polypeptide comprises cotton cellulose synthase
CC   CelA1.  Its amino acid sequence was deduced from a cDNA clone (see
CC   AAV60039) isolated from a cotton fibre cDNA library.  Expression of
CC   CelA1 and CelA2 (see AAW60040) in developing fibre is approximately
CC   500 times higher than the level of expression in other cotton
CC   tissues; approximately 1-2% of the 24 dpa mRNA is CelA mRNA.
CC   Sequence analysis of CelA1 and CelA2 showed regions of high homology
CC   (H-1, H-2 and H-3) between plant and bacteral proteins, as well as
CC   a region (P-CR) that is conserved among plant CelA genes, and
CC   another region (HVR) that is hypervariable between plant genes.
CC   Regions U-1 to U-4 are predicted to be critical for UDP-glc binding
CC   and catalysis in bacterial CelA proteins.  The isolated CelA cDNA
CC   clones are used in claimed methods for modifying fibre phenotype
CC   in a cotton plant and for modifying wood quality phenotype in a
CC   forest tree species.  Lignin can be reduced by over-expressing the
```

cellulose synthase product and diverting production into cellulose.

Sequence 974 AA;

Query Match 12.5%; Score 7; DB 19; Length 974;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

25 FVACHEC 31
||||||
25 FVACHEC 31

LSULT 15
AG30671
D AAG30671 standard; Protein; 2825 AA.
X
C AAG30671;
X
C 17-OCT-2000 (first entry)
X
C Arabidopsis thaliana protein fragment SEQ ID NO: 36709.
X
W Protein identification; signal transduction pathway; metabolic pathway;
W hybridisation assay; genetic mapping; gene expression control; promoter;
W termination sequence.
X
S Arabidopsis thaliana.
X
W EP1033405-A2.
X
D 06-SEP-2000.
X
F 25-FEB-2000; 2000EP-0301439.
X
R 25-FEB-1999; 99US-0121825.
R 05-MAR-1999; 99US-0123180.
R 09-MAR-1999; 99US-0123548.
R 23-MAR-1999; 99US-0125788.
R 25-MAR-1999; 99US-0126264.
R 29-MAR-1999; 99US-0126785.
R 01-APR-1999; 99US-0127462.
06-APR-1999; 99US-0128234.
08-APR-1999; 99US-0128714.
R 16-APR-1999; 99US-0129845.
19-APR-1999; 99US-0130077.
21-APR-1999; 99US-0130449.
R 23-APR-1999; 99US-0130510.
R 23-APR-1999; 99US-0130891.
R 28-APR-1999; 99US-0131449.
R 30-APR-1999; 99US-0132048.
R 30-APR-1999; 99US-0132407.
R 04-MAY-1999; 99US-0132484.
R 05-MAY-1999; 99US-0132485.
R 06-MAY-1999; 99US-0132486.
R 06-MAY-1999; 99US-0132487.
R 07-MAY-1999; 99US-0132863.
R 11-MAY-1999; 99US-0134256.
R 14-MAY-1999; 99US-0134218.
R 14-MAY-1999; 99US-0134219.
R 14-MAY-1999; 99US-0134221.
R 14-MAY-1999; 99US-0134370.
R 18-MAY-1999; 99US-0134768.
R 19-MAY-1999; 99US-0134941.
R 20-MAY-1999; 99US-0135124.
R 21-MAY-1999; 99US-0135353.
R 24-MAY-1999; 99US-0135629.
R 25-MAY-1999; 99US-0136021.
R 27-MAY-1999; 99US-0136392.
R 28-MAY-1999; 99US-0136782.
R 01-JUN-1999; 99US-0137222.
R 03-JUN-1999; 99US-0137528.
R 04-JUN-1999; 99US-0137502.

PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138543.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140053.
PR 23-JUN 1999; 99US-0140354.
PR 24-JUN 1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL 1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL 1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.

PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG 1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23 SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28 SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 12.5%; Score 7; DB 21; Length 2bases;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 44 IKEGRKV 50
|||||||
Db 205 IKEGRKV 211

Search completed: October 29, 2003, 18:34:29
Job time : 6.33269 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

GM nucleic - nucleic search, using sw model

Run on: October 30, 2003, 05:37:05 ; Search time 6759.03 Seconds
(without alignments)
12384.116 Million cell updates/sec

Title: US-09-980-043-5
Perfect score: 3444
Sequence: 1 gcggccgcggttaatcgccg.....gtgttgtagatagcggccgc 3444

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1486.6	43.2	3897	11	AY103655	AY103655 Zea mays
2	1485.8	43.1	3783	11	AY104730	AY104730 Zea mays
3	1402.2	40.7	3788	11	AY103701	AY103701 Zea mays
4	1209.2	35.1	3898	11	AY110415	AY110415 Zea mays
5	1160	33.7	3763	11	AY108113	AY108113 Zea mays
6	1125.4	32.7	3696	11	AY110079	AY110079 Zea mays
7	1123.8	32.6	2872	11	AY104236	AY104236 Zea mays
8	1102.4	32.0	3728	11	AY112236	AY112236 Zea mays
9	638	18.5	675	9	AV822205	AV822205 AV822205
c 10	593	17.2	593	9	AV782896	AV782896 AV782896
11	543	15.8	582	9	AV823575	AV823575 AV823575
c 12	500	14.5	637	9	AV524115	AV524115 AV524115
13	487.6	14.2	510	9	AV527222	AV527222 AV527222
14	483.8	14.0	799	14	CB894714	CB894714 EST647506
15	483.2	14.0	851	14	CB892280	CB892280 EST649249
16	480.8	14.0	792	14	CB671068	CB671068 OSJNBe04H
c 17	480	13.9	697	28	BH979329	BH979329 odf95b01.
18	454.8	13.2	953	14	CD439980	CD439980 EL01N0531
19	451.8	13.1	1189	11	AY107656	AY107656 Zea mays
20	451.6	13.1	715	13	BQ870049	BQ870049 QGD7P10.y
c 21	450	13.1	558	9	AV546832	AV546832 AV546832
22	446	13.0	788	10	BG584483	BG584483 EST486243
23	442.6	12.9	856	14	CB633016	CB633016 OSIIEb11J
24	440.6	12.8	862	14	CD437450	CD437450 EL01N0501
c 25	438.4	12.7	609	9	AV540720	AV540720 AV540720
26	437	12.7	738	12	BI920262	BI920262 EST540197
27	435.8	12.7	783	14	CB671503	CB671503 OSJNBe05B
c 28	427.4	12.4	430	9	AV784570	AV784570 AV784570
29	427	12.4	724	13	BQ875131	BQ875131 QGI7E03.y
30	426.4	12.4	791	10	BG351053	BG351053 103F12 Ma
c 31	425.4	12.4	580	9	AI994360	AI994360 701557762
32	423.6	12.3	779	14	CA916873	CA916873 EST641020
33	423	12.3	792	12	BI925066	BI925066 EST544955
34	418.6	12.2	714	14	CB629928	CB629928 OSIIEb06I
c 35	416.8	12.1	438	9	AB038726	AB038726 AB038726
36	416.8	12.1	439	9	AV525753	AV525753 AV525753
37	416.4	12.1	717	13	BQ867117	BQ867117 QGC9N09.y
c 38	416	12.1	528	14	W43509	W43509 22886 CD4-1
39	415.2	12.1	807	12	BI311236	BI311236 EST531298
40	413.6	12.0	701	13	BQ803259	BQ803259 WHE2835_E
41	413.6	12.0	717	13	BQ996712	BQ996712 QGG13J11.
42	412.4	12.0	530	13	BQ791285	BQ791285 E4162 Chi
43	411.4	11.9	731	13	BU007898	BU007898 QGH5M24.y
44	408.8	11.9	540	14	H36985	H36985 15114 Lambd
45	406.4	11.8	712	13	BU004334	BU004334 QGG4e11.y

ALIGNMENTS

RESULT 1
AY103655
LOCUS AY103655 3897 bp HTC 16-OCT-2002
DEFINITION Zea mays PCO096398 mRNA sequence.
ACCESSION AY103655
VERSION AY103655.1 GI:21206733
KEYWORDS HTC.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 3897)
AUTHORS Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitlatt,M.S., Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
TITLE Maize Mapping Project/DuPont Consensus Sequences for Design of Overgo Probes
JOURNAL Unpublished (2002)
REFERENCE 2 (bases 1 to 3897)
AUTHORS Coe,E.H.
TITLE Direct Submission
JOURNAL Submitted (25-APR-2002) Maize Mapping Project, University of Missouri, Columbia, MO 65211, USA
COMMENT If you are interested in getting corresponding physical clones, these are publicly available from [illegible] and may be found by BLAST searching at MSL, maizemap.org; ZmDB, [illegible].edu; TIGR, www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the

maize cDNA sequences is either Virginia Walbot, Stanford or Pat Schnable, Iowa State, then clones may be requested from ZmDB: www.zmdb.iastate.edu.

FEATURES Location/Qualifiers
source 1..3897
/organism="Zea mays"
/mol_type="mRNA"
/db_xref="MaizeDB:635496"
/db_xref="taxon:4577"
/clone_lib="Maize Mapping Project/DuPont Cornsensus library"
/note="this sequence is part of a project of EST assemblies resulting from the application of public contigs to seed DuPont contigs; this resource was assembled by DuPont as part of a collaboration for the overgo addressing of BACs in conjunction with the Maize Mapping Project"

BASE COUNT 983 a 631 c 1008 g 1068 t
ORIGIN

Query Match 43.2%; Score 1486.6; DB 11; Length 3897;
Best Local Similarity 67.3%; Pred. No. 0;
Matches 2187; Conservative 0; Mismatches 1034; Indels 30; Gaps 5;

```
Qy        9 GGTTA...CAACAGGAATGAGTTTGTCCTCATTAATGCCGATGAGAATG 68
Db      172 GGCTGGTGGCCGGCCCACAACCGCAACGAGCTCGTCGTCATCCGCCGCGACGGCGATC 231

Qy       69 CCCGAATAAGATGA...AGAGCTGAGTGGACAGACATGTCAAATCTGCAGAGATGAGA 128
Db      232 CCGGGCCGAAGCCG...GGAGCAGAACGGGCAGGTGTGCCAGATTTGCGGCGACGACG 291

Qy      129 TCGAAT...TGTTATGAAGAACCGTTTGTGGCATGTAACGAATGTGCATTCCCTGTGT 188
Db      292 TCGGCC...CCGGGAACCCTTCGTGGCGTGCAACGAGTGCGCCTTCCCCGTCT 351

Qy      189 GTAGAC...GTACGAAAGACGAGAAGGCAATCAAGCTTGTCCACAGTGCAAAA 248
Db      352 GCCGGG...GAGCGCCGGGAGGGCACGCAGAACTGCCCCCAGTGCAAGA 411

Qy      249 CCCGT...CAAACGT...AAGGAAGTCCAAGAGTTGAAGGTGATGAAGAGGAAGATGACA 308
Db      412 CTCGA...AAGCCGT...AAGGGCTGCCAACGTGTGACCGGTGACGAGGAGGAGGACGGCG 471

Qy      309 TTGA...AGTTTAGA...AATGAGTTTGA---GTATGGAAATAATGGGATTGGATTTGATCAGG 365
Db      472 TCGA...ACCTGGA...AACGAGTTCAACTGGGACGGCCATGACTCGCAGTCTGTGGCCGAGT 531

Qy      366 TTTCT...ATGTATGTCAA...TCTCGTCGCAACTCCGGTTTCCCACAATCTGATTTGGATT 425
Db      532 CCATG...TCTACGGCCAGA...AGCTACGGCCGTGGAGGTGACCCTAATGGCGCGCCACAAG 591

Qy      426 CAGCT...TCTGGCTCTT...ATTCCATTGCTGACTTACGGCGACGAGGACGTTGAGATTT 485
Db      592 CTTTC...AACCCTA...TGTTCCACTCCTCACCAACGGGCAAATGGTGGATGACATCC 651

Qy      486 CTTCT...GACAT...TTTATTGTTCCTCCTTCACTTGGTGGTCATGGCAATAGAGTTC 545
Db      652 CACCG...AGCAC...TGGTGCCTTCTTTCATGGGTGGTGGGGGAAAGAGGATAC 708

Qy      546 ATCCT...TTCTT...TCACCGTGGCTGCACATCGAAGGCTGATGGTACCTCAGA 605
Db      709 ATCCC...TTCT...AGCTTACCTGTGCAACCCAGGTCTATGGACCCATCCA 768

Qy      606 AAGAT...GGGTT...TTATGGAAGTGTCGCTTGGAAAGATCGGATGGAGGAATGGA 665
Db      769 AGGAT...CTGG...TATGGTAGTGTTGCTTGGAAGGAACGGATGGAGAATTGGA 828

Qy      666 AGAG...AAAG...AGAAACTTCAGGTTGTTAGGCATGAAGGAGATCCTGATTTTGAAG 725
Db      829 AGCAG...AAG...AGAGGATGCACCAGACGGGGA---------ATGATGGTGGTGGTGATG 879

Qy      726 ATGGT...ATGATGCTGATT...CCAATGATGGATGAGGGAAGGCAGCCATTGTCTATGAAGA 785
```

```
Db      880 ATGGTGACGATGCTGATCTACCACTAATGGATGAAGCAAGACAACAACTGTCCAGGAAAA 939

Qy      786 TACCAATCAAATCGAGCAAGATAAATCCTTACCGGATGTTAATTGTGCTACGTCTTGTGA 845
Db      940 TTCCACTTCCATCAAGCCAGATTAATCCATATAGGATGATTATCATTATTCGGCTTGTGG 999

Qy      846 TTCTTGGTCTCTTCTTTCACTACCGTATTCTTCACCCCGTCAAAGATGCATATGCTTTGT 905
Db     1000 TTTTGGGGTTCTTCTTCCACTACCGAGTGATGCATCCGGTGAATGATGCATTTGCTTTGT 1059

Qy      906 GGCTTATTTCTGTTATATGTGAGATATGGTTTGCTGTTTCATGGGTTCTTGATCAGTTCC 965
Db     1060 GGCGCATATCTGTTATCTGTGAAATCTGGTTTGCCATGTCTTGGATTCTTGATCAATTCC 1119

Qy      966 CTAAATGGTACCCTATCGAGCGAGAAACGTACTTGGACCGACTCTCATTAAGATATGAGA 1025
Db     1120 CAAAGTGGTTCCCTATTGAGAGAGAGACTTACCTAGACCGGCTGTCACTGAGGTTCGACA 1179

Qy     1026 AAGAAGGGAAACC...GGGACTATC...TGTGATGTATTTGTTAGTACAGTGGATCCAT 1085
Db     1180 AGGAAGGCCAGC...AAGTTG...AATTCTTTGTCAGTACGGTTGATCCCT 1239

Qy     1086 TGAAAGAGCCTCC...TTATTACTGCAAATAC...TTGTCTATTCTTGCTGTTGATTATC 1145
Db     1240 TAAAGGAACCTCCTTTGGTCACAACAAATACTGTTCTATCTATCCTTTCGGTGGATTATC 1299

Qy     1146 CTGTCGATAAGGTTGCTTGTTACGTATCTGATGATGGTGCTGCTATGCTTACTTTCGAAG 1205
Db     1300 CTGTTGATAAGGTTTCTTGCTATGTTTCTGATGATGGTGCTGCAATGCTAACGTTTGAAG 1359

Qy     1206 CTCTTTCTGAGACCGCTGAATTCGCAAGGAAATGGGTTCCTTTCTGCAAGAAATATTGTA 1265
Db     1360 CATTATCTGAAACATCTGAATTTGCAAAGAAATGGGTTCCTTTCTGCAAACGGTACAATA 1419

Qy     1266 TTGAGCCTCGTGCTCCCGAATGGTATTTCTGCCATAAAATGGACTACTTGAAGAATAAAG 1325
Db     1420 TTGAACCTCGCGCTCCAGAGTGGTACTTCCAACAGAAGATAGACTACTTGAAAGACAAGG 1479

Qy     1326 TTCATCCCGCATTTGTTAGGGAGCGGCGAGCCATGAAGAGAGATTATGAAGAATTCAAAG 1385
Db     1480 TGGCAGCAAACTTTGTTAGGGAGAGGAGAGCAATGAAGAGAGAGTATGAGGAATTCAAGG 1539

Qy     1386 TAAAGATCAATGCTTTAGTAGCAACAGCACAGAAAGTGCCTGAGGATGGTTGGACTATGC 1445
Db     1540 TGAGAATCAATGCCTTAGTTGCCAAAGCCCAGAAAGTTCCTGAAGAAGGATGGACAATGC 1599

Qy     1446 AAGACGGTACACCTTGGCCCGGTAATAGTGTGCGAGATCATCCTGGCATGATTCAGGTCT 1505
Db     1600 AAGATGGAACCCCCTGGCCTGGAAACAATGTTCGTGATCATCCTGGAATGATTCAGGTCT 1659

Qy     1506 TCCTTGGAAGTGACGGTGTTCGTGATGTCGAAAACAACGAGTTGCCTCGATTAGTTTACG 1565
Db     1660 TCCTTGGCCAAAGCGGAGGCCTTGACTGTGAGGGAAATGAACTGCCACGATTGGTTTATG 1719

Qy     1566 TTTCTCGTGAGAAGAGACCCGGATTTGATCACCATAAGAAGGCTGGAGCTATGAATTCCC 1625
Db     1720 TTTCTAGAGAGAAACGACCAGGCTATAACCATCATAAGAAAGCTGGTGCTATGAATGCAT 1779

Qy     1626 TGATACGAGTCTCTGGGGTTCTATCAAATGCTCCTTACCTTCTGAATGTCGATTGTGATC 1685
Db     1780 TGGTCCGAGTCTCTGCTGTACTAACAAATGCTCCATATTTGTTAAACTTGGATTGTGATC 1839

Qy     1686 ACTACATCAACAATAGCAAAGCTCTTAGAGAAGCAATGTGTTTCATGATGGATCCTCAGT 1745
Db     1840 ACTACATCAACAACAGCAAGGCTATAAAGGAAGCAATGTGTTTTATGATGGACCCTTTAC 1899

Qy     1746 CAGGAAAGAAAAATCTGTTATGTTCAGTTCCCTCAAAGGTTCGATGGGATTGATAGGCACG 1805
Db     1900 TAGGAAAGAAGGTTTGCTATGTACAGTTCCCTCAAAGATTTGATGGGATTGATCGCCATG 1959

Qy     1806 ATCGATACTCAAATCGCAATGTTGTGTTCTTTGATATCAATATGAAAGGTTTGGATGGGC 1865
```

```
Db       1960 ACCGATATGCTAACCGGAATGTTGTCTTTTTTGATATCAACATGAAAGGTTTGGATGGTA 2019

Qy       1866 TACAAGGGCCTATATACGTCGGTACAGGTTGTGTTTTCAGGAGGCAAGCGCTTTACGGAT 1925

Db       2020 TTCAGGGTCCAATTTATGTTGGTACTGGATGTGTATTTAGAAGGCAGGCATTATATGGTT 2079

Qy       1926 TTGATGCACCGAAGAAGAAGAAGGGCCCACGTAAGACATGCAATTGCTGGCCAAAATGGT 1985

Db       2080 ATGATGCCCCCAAAACAAAGAAGCCACCATCAAGGACTTGCAACTGCTGGCCCAAGTGGT 2139

Qy       1986 GTCTCCTATGTTTTGGTTCAAGAAAGAATCGTAAAGCAAAGACAGTGGCTGCGGATAAGA 2045

Db       2140 GCTTTTGCTGTTGCTGCTTTGGCAATAGGAAGCAAAAGAAGACTACCAAACCCAAAACAG 2199

Qy       2046 AGAAGAAGAA----------TAGGGAAGCGTCAAAGCAGATCCACGCATTAGAAAATA 2093

Db       2200 AGAAGAAAAAGTTATTATTTTTCAAGAAAGAAGAGAACCAATCCCCTGCATATGCTCTTG 2259

Qy       2094 TCGAAGAGGGCCGCGGTCATAAAGTTCTTAACGTAGAACAGTCAACCGAGGCAATGCAAA 2153

Db       2260 GTGAAATTGACGAAGCTGCTCCAGGAGCTGAGAATGAAAAGGCCGGTATTGTAAATCAAC 2319

Qy       2154 TGAAGTTGCAGAAGAAATATGGGCAGTCTCCTGTATTTGTTGCATCTGCGCGTCTGGAGA 2213

Db       2320 AAAAATTAGAAAAGAAATTTGGCCAATCTTCTGTTTTTGTTACATCCACACTTCTCGAGA 2379

Qy       2214 ATGGTGGGATGGCTAGAAACGCAAGCCCGGCTTGTCTGCTTAAAGAAGCCATCCAAGTCA 2273

Db       2380 ATGGTGGAACCTTGAAGAGTGCAAGTCCTGCTTCTCTTTTGAAAGAAGCTATACATGTCA 2439

Qy       2274 TTAGTCGCGGATATGAAGATAAAACTGAATGGGGAAAAGAGATTGGGTGGATCTATGGTT 2333

Db       2440 TTAGTTGTGGTTATGAAGACAAGACAGACTGGGGAAAAGAGATTGGCTGGATCTATGGAT 2499

Qy       2334 CTGTTACCGAAGATATTCTTACGGGTTCTAAGATGCATTCTCATGGTTGGAGACATGTTT 2393

Db       2500 CAGTTACAGAAGATATTCTAACTGGTTTCAAGATGCATTGTCATGGTTGGCGGTCAATTT 2559

Qy       2394 ATTGTACACCAAAGTTAGCGGCTTTCAAAGGATCAGCTCCAATCAATCTTTCGGATCGTC 2453

Db       2560 ACTGCATACCTAAACGGGTTGCATTCAAAGGTTCTGCACCTCTGAATCTTTCAGATCGTC 2619

Qy       2454 TCCATCAAGTTCTTCGATGGGCGTTTGAGTCGGTTGAGATTTCTTGAGTAGGCATTGTC 2513

Db       2620 TTCACCAGGTGCTTCGGTGGGCTTTGGGTCTATTGAGATCTTCTTCAGCAATCATTGCC 2679

Qy       2514 CTATTTGGTATGGTTATGGAGGTGGGTTGAAATGGCTTGAGCGGTTGTCCTACATTAACT 2573

Db       2680 CTCTTTGGTATGGGTATGGTGGCGGTCTGAAATTTTTGGAAAGATTTTCCTACATCAACT 2739

Qy       2574 CTGTGGTTTACCCGTGGACCTCTCTACCGCTCATCGTTTACTGTTCTCTCCCTGCCATCT 2633

Db       2740 CCATCGTGTATCCTTGGACATCTATTCCCCTCTTGGCTTACTGTACATTGCCTGCCATCT 2799

Qy       2634 GTCTTCTCACTGGAAAATTCATCGTTCCCGAGATTAGCAACTATGCGAGTATCCTCTTCA 2693

Db       2800 GTTTATTGACAGGGAAATTTATCACTCCAGAGCTGAATAATGTTGCCAGCCTGTGGTTCA 2859

Qy       2694 TGGCGCTCTTCTCGTCGATTGCAATAACGGGTATTCTCGAGATGCAATGGGGCAAAGTTG 2753

Db       2860 TGTCACTTTTTATCTGCATTTTTGCTACGAGCATCCTAGAAATGAGATGGAGTGGTGTTG 2919

Qy       2754 GGATCGATGATTGGTGGAGAAACGAACAGTTTTGGGTCATTGGAGGTGTTTCTGCGCATC 2813

Db       2920 GAATTGATGACTGGTGGAGGAATGAGCAGTTCTGGGTCATTGGAGGTGTGTCCTCACACC 2979

Qy       2814 TGTTTGCTCTCTTCCAAGGTCTCCTCAAGGTTCTTGCTGGTGTCGACACTAACTTCACAG 2873

Db       2980 TCTTTGCTGTGTTCCAGGGACTTCTCAAGGTCATAGCTGGTGATGATACAAGCTTCACCG 3039

Qy       2874 TCACATCAAAAGCAGCTGATGATGGAGAGTTCTCTGACCTTTACCTCTTCAAATGGACTT 2933

Db       3040 TGACATCAAAGGGTGGAGATGATGAGGAGTTCTCAGAGCTATATACATTCAAATGGACTA 3099

Qy       2934 CACTTCTCATCCCTCCAATGACTCTACTCATCATAAACGTCATTTT  ATAGTCGGAG 2993

Db       3100 CCTTATTGATACCTCCTACCACCTTGCTTCTATTGAACTTG TGTCGCTGGCG 3159

Qy       2994 TCTCTGATGCCATCAGCAATGGATACGACTCGT  GACTGTTCT 3053

Db       3160 TTTCAAATGCGATCAATAACGGATATGAGTCAT  AAGCTATTCT 3219

Qy       3054 TTGCACTTTGGGTCATCATTCATCTTTACCCGT  GGAAACAAG 3113

Db       3220 TTGCATTTTGGGTGATTGTCCATCTTTATCCGTT  GGAAGGCAAA 3279

Qy       3114 ATAGAATGCCAACCATTATTGTCGTCTGGTC  TTACACTTC 3173

Db       3280 ACAGGACACCAACGATTGTCATCGTCTGGTCCAT  TTCTCGCTCC 3339

Qy       3174 TTTGGGTCCGGGTTAATCCGTTTGTGGCGA---  AGATCTGTG 3230

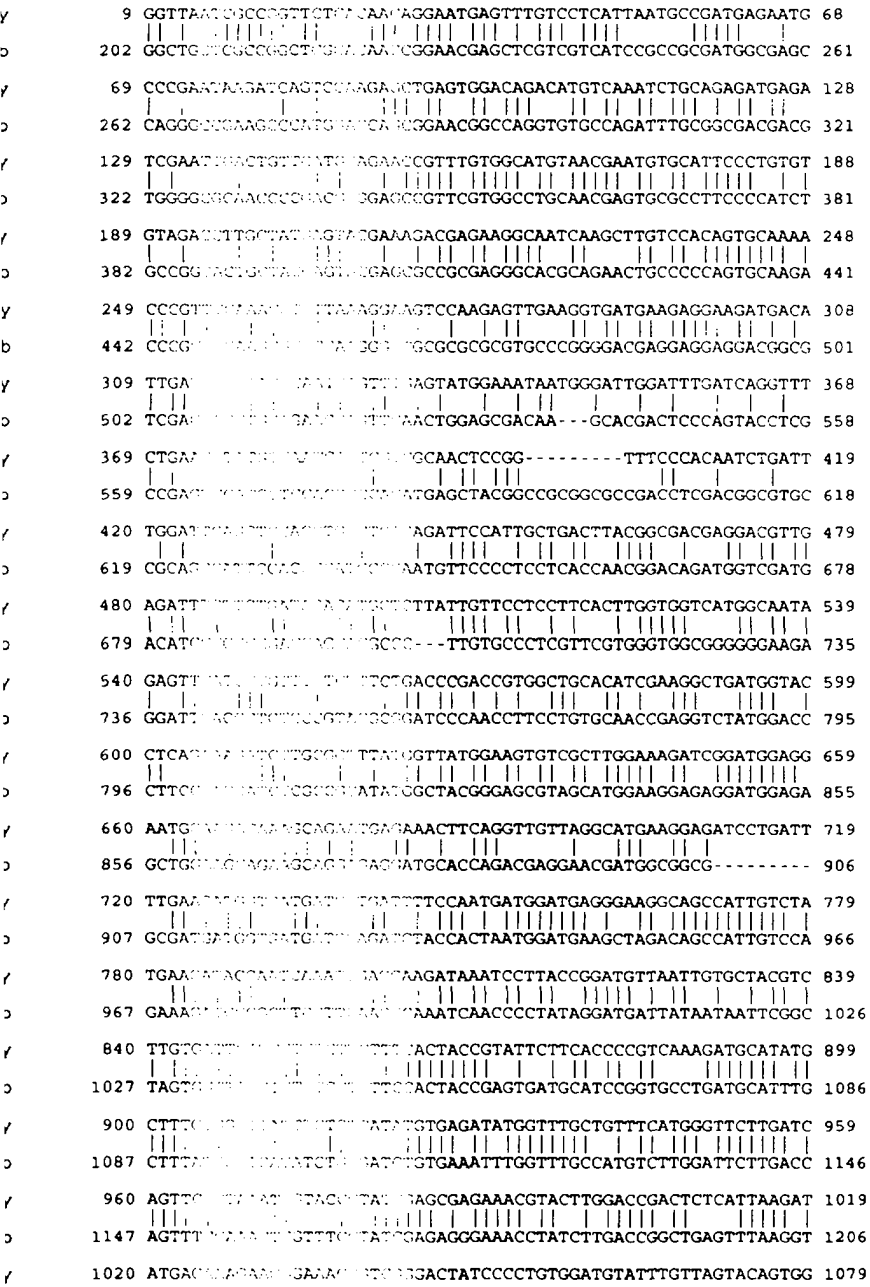
Db       3340 TTTGGGTTCGGATTGATCCTTTCCTTGCGAAG  AGGAGTGTG 3399

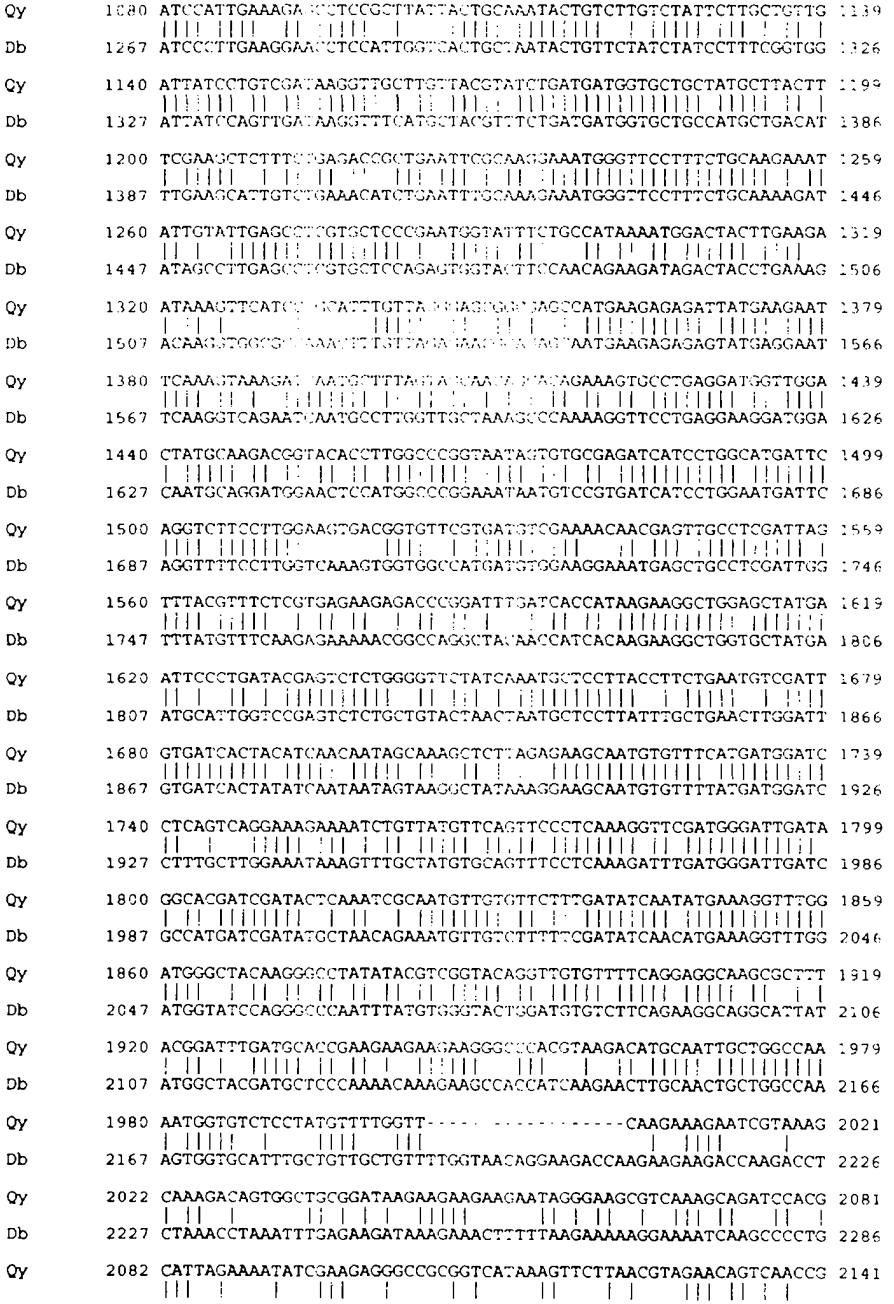
Qy       3231 GTTTAGACTGC 3241

Db       3400 GTTTGGATTGC 3410
```

```
RESULT 2
AY104730
LOCUS       AY104730               3783 bp    RNA     linear   PTC 16-OCT-2002
DEFINITION  Zea mays PCO100501 mRNA sequence.
ACCESSION   AY104730
VERSION     AY104730.1  GI:21207808
KEYWORDS    HTC.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 3783)
  AUTHORS   Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Wright,M.S.,
            Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.
  TITLE     Maize Mapping Project/DuPont Consensus Sequence Design of
            Overgo Probes
  JOURNAL   Unpublished (2002)
REFERENCE   2  (bases 1 to 3783)
  AUTHORS   Coe,E.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-APR-2002) Maize Mapping Project, University of
            Missouri, Columbia, MO 65211, USA
COMMENT     If you are interested in getting corresponding physical clones,
            these are publicly available from ZmDB and may be found by BLAST
            searching at MSL, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR,
            www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. While the source of the
            maize cDNA sequences is either Virginia Walbot, Stanford or Pat
            Schnable, Iowa State, then clones may be requested from ZmDB:
            www.zmdb.iastate.edu.
FEATURES             Location/Qualifiers
     source          1..3783
                     /organism="Zea mays"
                     /mol_type="mRNA"
                     /db_xref="MaizeDB:635765"
                     /db_xref="taxon:4577"
                     /clone_lib="Maize Mapping Project/DuPont Consensus
                     Library"
                     /note="this sequence is part of a project of EST
                     assemblies resulting from the application of public
                     contigs to seed DuPont contigs; this sequence was
                     assembled by DuPont as part of a collaboration for the
                     overgo addressing of BACs in conjunction with the Maize
                     Mapping Project"
BASE COUNT      933 a    908 c   1002 g    940 t
ORIGIN
```

Query Match 43.1%; Score 1485.8; DB 11; Length 3783;
Best Local Similarity 67.6%; Pred. No. 0;
Matches 2201; Conservative 0; Mismatches 1017; Indels 45; Gaps 6;





```
Db       2287 CATATGCTCTTGGTGAAATTGATGAAGCTGCTCCAGGAGCTGAAAATGAAAAGGCTAGTA 2346

Qy       2142 AGGCAATGCAAATGAAGTTGCAGAAGAAATATGGGCAGTCTCCTGTATTTGTTGCATCTG 2201
             | ||  |||  ||||||| | |||||| ||| ||||||  | ||  ||||||||||||
Db       2347 TTGTAAATCAACAGAAGTTGGAAAAGAAATTTGGCCAGTCTTCAGTTTTTGTTGCATCCA 2406

Qy       2202 CGCGTCTGGAGAATGGTGGGATGGCTAGAAACGCAAGCCCGGCTTGTCTGCTTAAAGAAG 2261
             | | ||| |||||||||||| |      |  |  || || || |||| ||| || || ||||
Db       2407 CACTTCTTGAGAATGGTGGAACCCTGAAGAGTGCCAGTCCAGCTTCTCTTCTGAAGGAAG 2466

Qy       2262 CCATCCAAGTCATTAGTCGCGGATATGAAGATAAAACTGAATGGGGAAAAGAGATTGGGT 2321
             | || || ||||| ||  | ||.|||||||| ||||| |  ||||||||||| ||||| |
Db       2467 CTATACATGTCATCAGCTGTGGATATGAAGACAAAACAGGCTGGGGAAAAGATATTGGTT 2526

Qy       2322 GGATCTATGGTTCTGTTACCGAAGATATTCTTACGGGTTCTAAGATGCATTCTCATGGTT 2381
             |||| ||||| || || || ||||||||||||||| || | |||||||| |  |||||||
Db       2527 GGATTTATGGATCAGTCACAGAAGATATTCTTACTGGGTTTAAGATGCACTGCCATGGTT 2586

Qy       2382 GGAGACATGTTTATTGTACACCAAAGTTAGCGGCTTTCAAAGGATCAGCTCCAATCAATC 2441
             ||  |    |||| || | ||| ||     || || |||||||| || || ||  ||||||
Db       2587 GGCGGTCAATTTACTGCATACCTAAACGGGCCGCCTTCAAAGGTTCCGCACCTCTCAATC 2646

Qy       2442 TTTCGGATCGTCTCCATCAAGTTCTTCGATGGGCGCTTGGGTCGGTTGAGATTTTCTTGA 2501
             |||| |||||||| || || |||||||| ||||| ||||| ||  |||| ||||||||  |
Db       2647 TTTCCGATCGTCTTCACCAGGTTCTTCGGTGGGCTCTTGGTTCAATTGAAATTTTCTTCA 2706

Qy       2502 GTAGGCATTGTCCTATTTGGTATGGTTATGGAGGTGGGTTGAAATGGCTTGAGCGGTTGT 2561
             | |  || || ||| | ||||| ||||| ||||  | || |  || ||  |||| |
Db       2707 GCAACCACTGCCCTCTCTGGTATGGGTATGGTGGTGGACTAAAGTTCCTGGAAAGGTTTT 2766

Qy       2562 CCTACATTAACTCTGTGGTTTACCCGTGGACCTCTCTACCGCTCATCGTTTACTGTTCTC 2621
             | ||||||||||  | || |||| |||| ||| | |||||| | |  || ||  |
Db       2767 CGTACATTAACTCCATCGTATACCCTTGGACATCTATCCCGCTCTTGGCCTATTGCACAT 2826

Qy       2622 TCCCTGCCATCTGTCTTCTCACTGGAAAATTCATCGTTCCCGAGATTAGCAACTATGCGA 2681
             | ||||||||||  | || || || |||||  ||  || ||| ||| |||   ||| |
Db       2827 TGCCTGCCATCTGCTTGCTGACAGGGAAATTTATCACGCCAGAGCTTAACAATGTTGCCA 2886

Qy       2682 GTATCCTCTTCATGGCGCTCTTCTCGTCGATTGCAATAACGGGTATTCTCGAGATGCAAT 2741
             |  |    |||||| | || |||    |  |||       ||| | || || || |||  ||
Db       2887 GCCTGTGGTTCATGTCACTTTTCATCTGCATTTTTGCTACGAGCATCCTGGAAATGAGAT 2946

Qy       2742 GGGGCAAAGTTGGGATCGATGATTGGTGGAGAAACGAACAGTTTTGGGTCATTGGAGGTG 2801
             || |    || || ||||||| ||||| ||||||| ||||||||||||||||||| |
Db       2947 GGAGTGGTGTAGGCATCGATGACTGGTGGAAAACGAGCAGTTTTGGGTCATTGGAGGCG 3006

Qy       2802 TTTCTGCGCATCTGTTTGCTCT[illegible]TCCAA[illegible]TCTTGCTGGTGTCGACA 2861
             | ||| | ||||| ||||||   ||| || || ||||||  | ||||||| |||
Db       3007 TGTCTTCACATCTCTTTGCTGTG[illegible]TCCAG[illegible]GACTCCTCAAG[illegible]TCATAGCTGGTGTAGACA 3066

Qy       2862 CTAACTTCACAGTCACATCAAAAGCAGCTGATGATGGAGAGTTCTCTGACCTTTACCTCT 2921
             | | |||||| || ||||| || |  |  || || |  |||||| || || |||   |
Db       3067 CGAGCTTCACTGTGACATCCAAGGGCGGAGACGACGAGGAGTTCTCAGAGCTGTACACAT 3126

Qy       2922 TCAAATGGACTTCACTTCTCATCCCTCCAATGACTCTACTCATCATAAACGTCATTGGAG 2981
             |||||||||  | |||| || |||| |  || || ||| |  | ||  |||||||||
Db       3127 TCAAATGGACGACCCTTCTGATACCTCCGACAACCCTGCTCCTACTGAACTTCATTGGAG 3186

Qy       2982 TCATAGTCGGAGTCTCTGATGCCATCAGCAATGGATACGACTCGTGGGGACCGCTTTTCG 3041
             \  |||  ||  |||| |||| |||| ||| |||| || || |||| || || |||| |||
Db       3187 TGGTAGCTGGCATCTCCAATGCGATCAACAACGGATATGAATCATGGGGCCCCCTGTTCG 3246

Qy       3042 GAAGACTGTTCTTTGCACTTTGGGTCATCATTCATCTTTACCCGTTCCTTAAAGGTTTGC 3101
             | |  || |||||||| |||||| ||| | ||||||||||||||| || ||| ||
Db       3247 GGAAGCTCTTCTTTGCATTTTGGGTGATCGTCCATCTTTACCCGTTCCTCAAGGGTCTGG 3306

Qy       3102 TTGGGAAACAAGATAGAATGCCAACCATTATTGTCGTCTGGTCCATCCTCCTGGCCTCGA 3161
             |||.||  ||  | || | |||||| || |  | |||||||||||||:|||| ||||
Db       3307 TTGGGAGGCAGAACAGGACGCCAACGATTGTCATTGTCTGGTCCATCCTCCTGGCTTCGA 3366

Qy       3162 TTCTTACACTTCTTTGGGTCCGGGTTAATCCGTTTGTGGCGA---AAGGCGGTCCTATTC 3218
             |  |  | ||||||||||.| |  | ||||  | ||||   | |  ||||  :
Db       3367 TCTTCTCGCTGCTTTGGGTCCGGATCGACCCGTTCCTTGCGAAGGATGATGGTCCCCTGT 3426

Qy       3219 TCGAGATCTGTGGTTTAGACTGC 3241
             | |||   |||||| | || |||
Db       3427 TGGAGGAGTGTGGTCTGGATTGC 3449


RESULT 3
AY103701
LOCUS       AY103701               3788 bp    [illegible]    [illegible]  16-OCT-2002
DEFINITION  Zea mays PCO120363 mRNA sequence.
ACCESSION   AY103701
VERSION     AY103701.1  GI:21206779
KEYWORDS    HTC.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Strepto[illegible]; [illegible]racheophyta;
            Spermatophyta; Magnoliophyta; Lil[illegible]eae; PACCAD
            clade; Panicoideae; Andropogoneae; [illegible]
REFERENCE   1  (bases 1 to 3788)
  AUTHORS   Hainey,C.F., Dolan,M., Miao,G.H., [illegible],M.S.,
            Arthur,L.W., Hanafey,M., Morgante,[illegible]
  TITLE     Maize Mapping Project/DuPont Conse[illegible]esign of
            Overgo Probes
  JOURNAL   Unpublished (2002)
REFERENCE   2  (bases 1 to 3788)
  AUTHORS   Coe,E.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-APR-2002) Maize Mapp[illegible], [illegible]ity of
            Missouri, Columbia, MO 65211, USA
COMMENT     If you are interested in getting c[illegible]l clones,
            these are publicly available from [illegible]d by BLAST
            searching at MSL, maizemap.org; Zm[illegible].edu; TIGR,
            www.tigr.org; or NCBI, www.ncbi.nl[illegible]source of the
            maize cDNA sequences is either Vi[illegible]rd or Pat
            Schnable, Iowa State, then clones [illegible]m ZmDB:
            www.zmdb.iastate.edu.
FEATURES             Location/Qualifiers
     source          1. .3788
                     /organism="Zea mays"
                     /mol_type="mRNA"
                     /db_xref="MaizeDB:637035"
                     /db_xref="taxon:4577"
                     /clone_lib="Maize Mappin[illegible]sensus
                     Library"
                     /note="this sequence is [illegible]ST
                     assemblies resulting fro[illegible]public
                     contigs to seed DuPont c[illegible]was
                     assembled by DuPont as p[illegible]n for the
                     overgo addressing of BAC[illegible]n the Maize
                     Mapping Project"
BASE COUNT      851 a    954 c   1114 g    866 [illegible]    3 [illegible]
ORIGIN

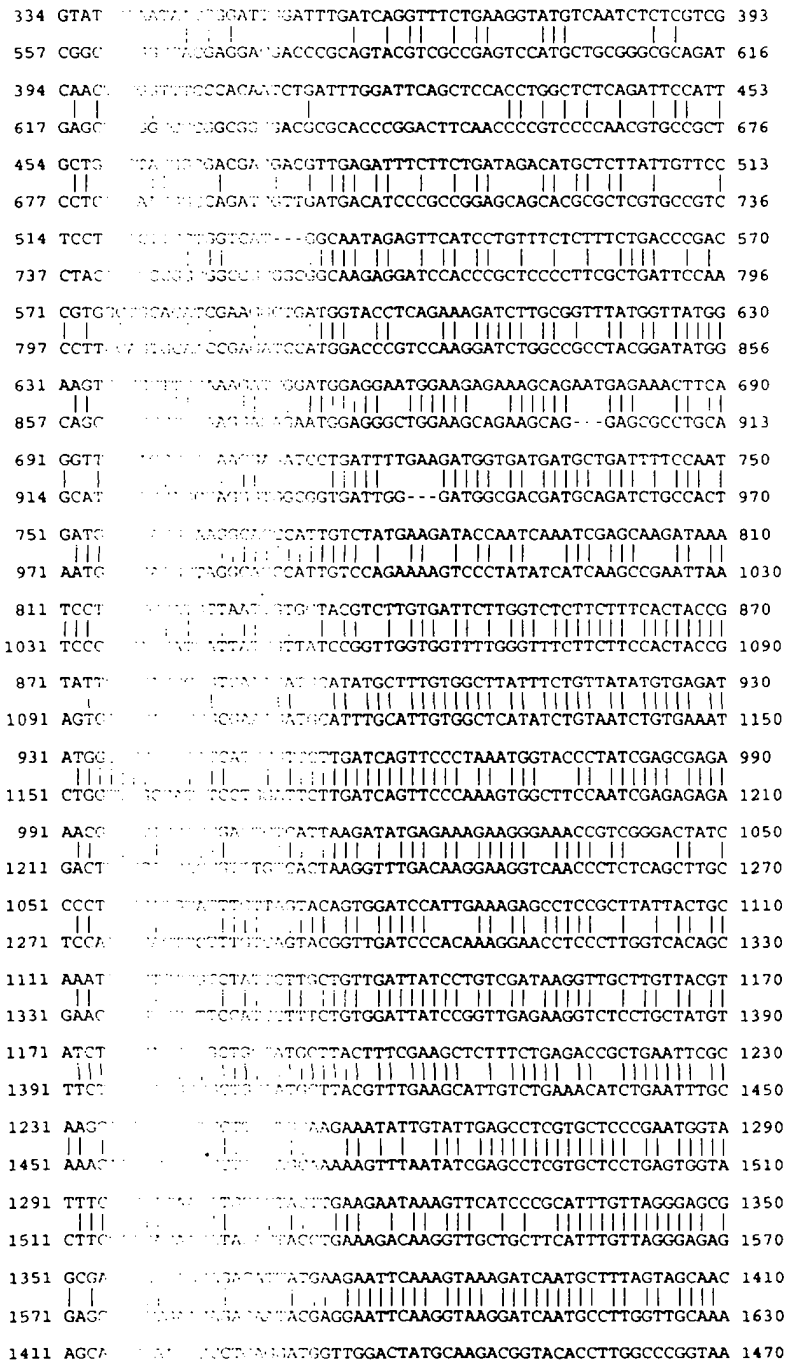
  Query Match           40.7%;  Score 1402.2;  DB 11;  [illegible]  788;
  Best Local Similarity   66.6%;  Pred. No. 0,
  Matches 2123;  Conservative    0;  Mismatche[illegible]  Indels    4;  Gaps    7;

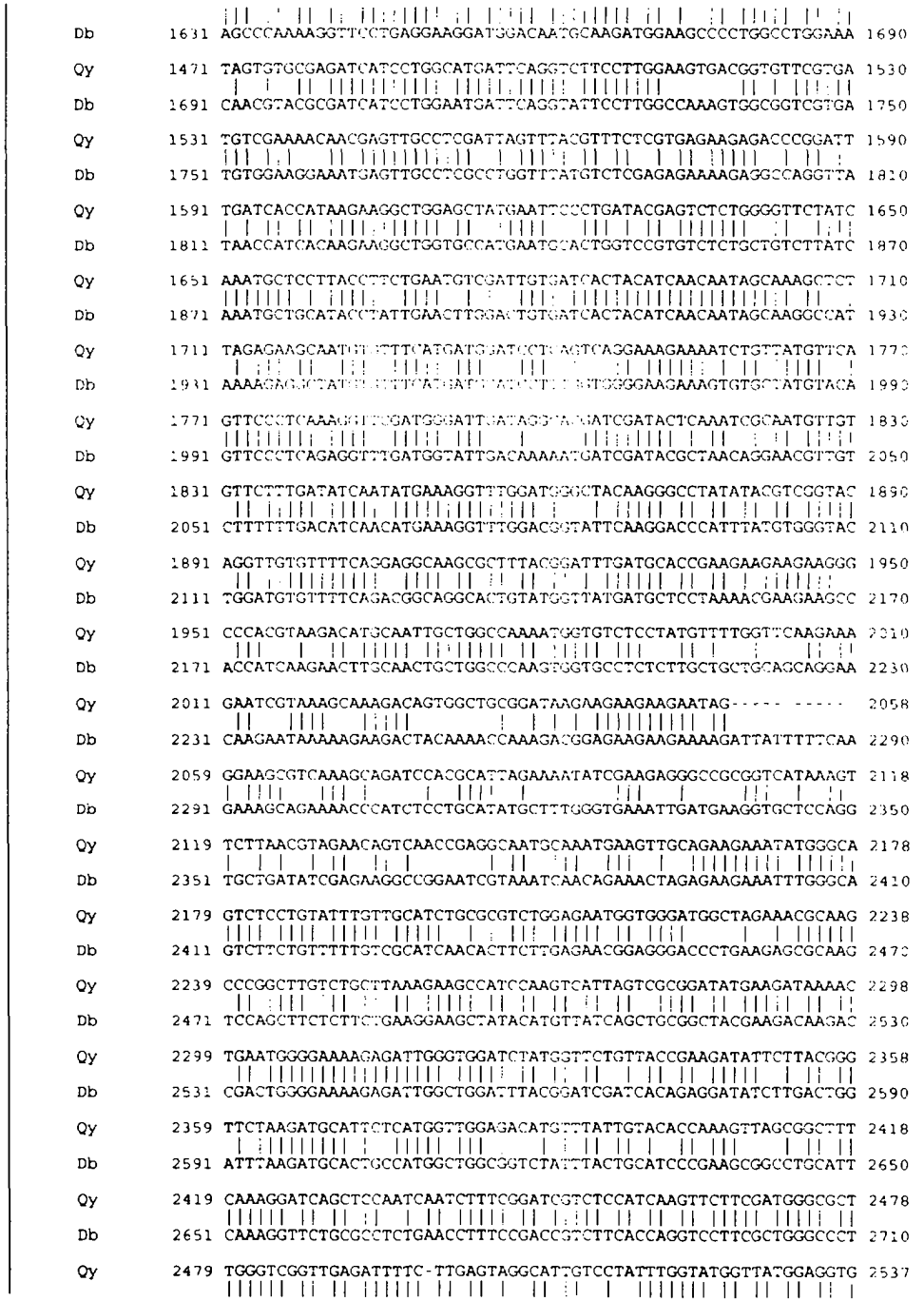
Qy       103 GACATGTCAAATCTGCAGAGATGAGATCGAA[illegible]GACTGTTGA[illegible]AA[illegible]TTTGTGGC 162
             | | || || || ||| | || ||| |||   |[illegible]  | ||[illegible]  ||  |||||
Db       317 GCCGTGCCAGATATGCGGCGACGAGGTCGGGG[illegible]GGCTTCGAC[illegible]AC[illegible]TTCGTGGC 376

Qy       163 ATGTAACGAATGTGCATTCCCTGTGTGTAGA[illegible]TGCTATGA[illegible]AC[illegible]ACGAGAAGG 222
             || ||||| || || ||||| || ||  |    |[illegible]  |[illegible]  || ||
Db       377 GTGCAACGAGTGCGCCTTCCCCGTCTGCCGCC[illegible]ACGA[illegible]AC[illegible]CGCGAGGG 436

Qy       223 CAATCAAGCTTGTCCACAGTGCAAAACCCGTT[illegible]TTCTTAA[illegible]CCAAGAGT 282
             |   || || || || |||||||  ||||| . ||;    || |  |  ||  | ||
Db       437 CTCGCAGGCGTGCCCGCAGTGCAGGACCCGCT[illegible]CGTTAA[illegible]CCGCGGGT 496

Qy       283 TGAAGGTGATGAAGAGGAAGATGACATTGATC[illegible]------TGA 333
             |  || || || ||||| || | | | ||                    ||
Db       497 GGCTGGGGACGAGGAGGAGGACGGCGTCGAC[illegible]CTGCAGGA 556
```





```
Db       2711 TGGGTCCGTCGAAATTTTCTTTCAGNAANCAATGGNCACTTTGGTACGGATACGGCGGCG 2770

Qy       2538 GGTTGAAATGGCTTGAGCGGTTGTCCTACATTAACTCTGTGGTTTACCCGTGGACCTCTC 2597

Db       2771 GGCTAAAATTCCTGGAAAGGTTTTCTTATATCAACTCCATCGTTTATCCCTGGACGTCCA 2830

Qy       2598 TACCGCTCATCGTTTACTGTTCTCTCCCTGCCATCTGTCTTCTCACTGGAAAATTCATCG 2657

Db       2831 TTCCTCTCCTGGCTTACTGTACCTTGCCTGCCATCTGCCTGCTCACGGGGAAGTTTATCA 2890

Qy       2658 TTCCCGAGATTAGCAACTATGCGAGTATCCTCTTCATGGCGCTCTTCTCGTCGATTGCAA 2717

Db       2891 CACCAGAGCTTACCAATGTCGCCAGTATCTGGTTCATGGCACTTTTCATCTGCATCTCCG 2950

Qy       2718 TAACGGGTATTCTCGAGATGCAATGGGGCAAAGTTGGGATCGATGATTGGTGGAGAAACG 2777

Db       2951 TGACCGGCATCCTGGAAATGAGGTGGAGTGGCGTGGCCATCGACGACTGGTGGAGGAACG 3010

Qy       2778 AACAGTTTTGGGTCATTGGAGGTGTTTCTGCGCATCTGTTTGCTCTCTTCCAAGGTCTCC 2837

Db       3011 AGCAGTTCTGGGTCATCGGAGGCGTTTCGGCGCATCTGTTCGCGGTGTTCCAGGGCCTGC 3070

Qy       2838 TCAAGGTTCTTGCTGGTGTCGACACTAACTTCACAGTCACATCAAAAGCAGCTGATGATG 2897

Db       3071 TGAAGGTGTTCGCCGGCATCGACACGAGCTTCACCGTGACGTCGAAGGCCGGGGACGACG 3130

Qy       2898 GAGAGTTCTCTGACCTTTACCTCTTCAAATGGACTTCACTTCTCATCCCTCCAATGACTC 2957

Db       3131 AGGAGTTCTCGGAGCTGTACACGTTCAAGTGGACCACCCTGCTGATACCCCCGACCACGC 3190

Qy       2958 TACTCATCATAAACGTCATTGGAGTCATAGTCGGAGTCTCTGATGCCATCAGCAATGGAT 3017

Db       3191 TCCTCCTGCTGAACTTCATCGGGGTGGTGGCCGGGATCTCGAACGCGATCAACAACGGGT 3250

Qy       3018 ACGACTCGTGGGGACCGCTTTTCGGAAGACTGTTCTTTGCACTTTGGGTCATCATTCATC 3077

Db       3251 ACGAGTCGTGGGGCCCCCTGTTCGGGAAGCTCTTCTTCGCGTTCTGGGTGATCGTCCACC 3310

Qy       3078 TTTACCCGTTCCTTAAAGGTTTGCTTGGGAAACAAGATAGAATGCCAACCATTATTGTCG 3137

Db       3311 TGTACCCGTTCCTCAAGGGTCTGGTGGGGAGGCAGAACAGGACGCCGACGATCGTCATCG 3370

Qv       3138 TCTGGTCCATCCTCCTGGCCTCGATTCTTACACTTCTTTGGGTCCGGGTTAATCCGTTTG 3197

Db       3371 TCTGGTCCATCCTGCTGGCCTCGATCTTCTCGCTCCTGTGGGTCCGCGTCGACCCGTTCC 3430

Qy       3198 TGGC---GAAAGGCGGTCCTATTC[illegible]ACTGCTTGTGATTCGATT 3254

Db       3431 TCGCCAAGAGCGACGGCCCGCTC[illegible]GGAGGAGTGTGGCCTGAACTGCAACTGAAGTGGGG 3490

Qy       3255 GACCGGTG 3262

Db       3491 GCCCCCTG 3498
```

```
RESULT 4
AY110415
LOCUS       AY110415                3898 bp    mRNA    linear   HTC 17-OCT-2002
DEFINITION  Zea mays CL1166_1 mRNA sequence.
ACCESSION   AY110415
VERSION     AY110415.1  GI:21214824
KEYWORDS    HTC.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 3898)
  AUTHORS   Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S.,
            Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
  TITLE     Maize Mapping Project/DuPont Consensus Sequences for Design of
            Overgo Probes
  JOURNAL   Unpublished (2002)
REFERENCE   2  (bases 1 to 3898)
  AUTHORS   Coe,E.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-APR-2002) Maize Mapp[illegible]versity of
            Missouri, Columbia, MO 65211, USA
COMMENT     If you are interested in getting c[illegible]ing [illegible]ical clones,
            these are publicly available from [illegible]ay [illegible]ound by BLAST
            searching at MSL, maizemap.org; Zm[illegible]mdb[illegible]ate.edu; TIGR,
            www.tigr.org; or NCBI, www.ncbi.nl[illegible]. W[illegible]he source of the
            maize cDNA sequences is either Vir[illegible]bot, [illegible]nford or Pat
            Schnable, Iowa State, then clones [illegible]pe.[illegible] from ZmDB:
            www.zmdb.iastate.edu.
FEATURES             Location/Qualifiers
     source          1..3898
                     /organism="Zea mays"
                     /mol_type="mRNA"
                     /db_xref="MaizeDB:630057"
                     /db_xref="taxon:4577"
                     /clone_lib="Maize Mapping [illegible]prnsensus
                     Library"
                     /note="this sequence is p[illegible]of EST
                     assemblies resulting from [illegible]of public
                     contigs to seed DuPont co[illegible]rce was
                     assembled by DuPont as pa[illegible]tion for the
                     overgo addressing of BACs [illegible]ith the Maize
                     Mapping Project"
BASE COUNT      932 a    841 c    991 g    1054 [illegible]
ORIGIN

  Query Match           35.1%;  Score 1208.2[illegible]  L[illegible]  3898;
  Best Local Similarity   62.4%;  Pred. No. 5.6[illegible]
  Matches 1995;  Conservative    0;  Mismatches [illegible];  Ind[illegible]  48;  Gaps    8;

Qy           7 GCGGTTAATCGCCGGTTCTCACAACAGGAATGA[illegible]TCA[illegible]-----TAATGCCGA 60

Db         203 GGGGATGGTGGCGGGCTCGCACAACCGCAACGA[illegible]CATGAT[illegible]CACGACGGCGA 262

Qy          61 TGAGAATGCCCGAATAAGATCAGTCCAAGAGCTGA[illegible]CAAATCTGCAG 120

Db         263 CGCGCCGGGCCCGGCTAAGCCCACGAAGAGTGC[illegible]CAGATTTGCGG 322

Qy         121 AGATGAGATCGAATTGACTGTTGATGGAGAACC[illegible]GAATGTGCATT 180

Db         323 CGACTCTGTGGGCGTTTCAGCCACTGGTGATGT[illegible]GAGTGTGCCTT 382

Qy         181 CCCTGTGTGTAGACCTTGCTATGAGTACGAAA[illegible]GCTTGTCCACA 240

Db         383 CCCTGTCTGCCGCCCTTGCTATGAGTACGAGC[illegible]GCTGCCCTCA 442

Qy         241 GTGCAAAACCCGTTTCAAACGTCTTAAAGGAA[illegible]ATGAAGAGGA 300

Db         443 GTGCAAGACTAGATACAAGAGACAGAAAGGTA[illegible]GATGAGGATGA 502

Qy         301 AGATGACATTGATGATTTAGACAATGAGTTTGA[illegible]ATTGGATTTGA 360

Db         503 GGAAGATGTTGATGACCTAGACAATGAATTCAA[illegible]GTGGGAAGGGC 560

Qy         361 TCAGGTTTCTGAAGGTATGTCAATCTCTCGTCG[illegible]CAATCTGATTT 420

Db         561 CCAGAGTGGCAGCTGCAAGGAGATGACGC------ --------[illegible]ATCTGTCTTCATC 604

Qy         421 GGATTCAGCTCCACCTGGCTCTCAGATTCCATTG[illegible]ACGAGGACGTTGA 480

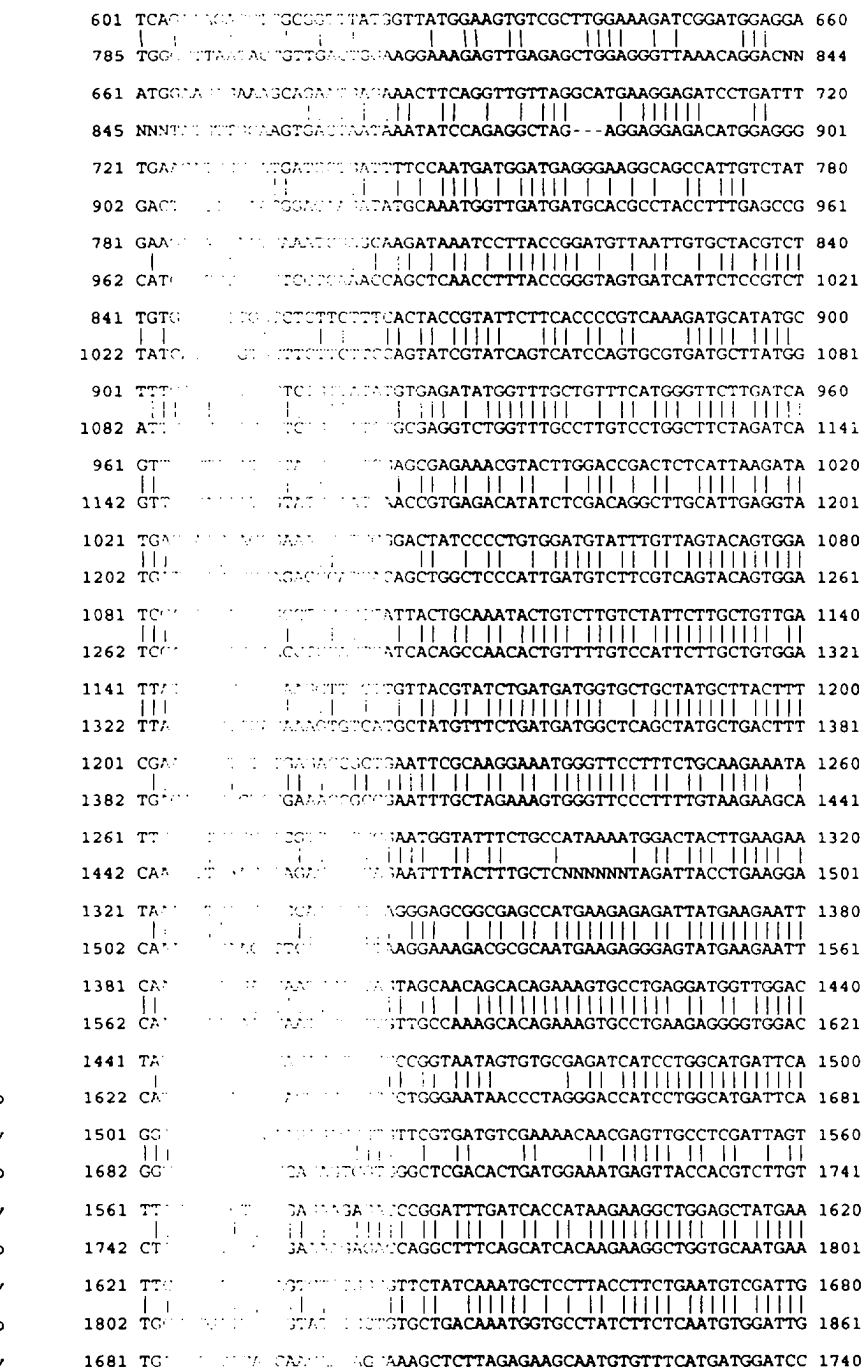
Db         605 TGCTCGCCATGACCCACACCATCGGATTCCACGC[illegible]CAGATATCTGG 664

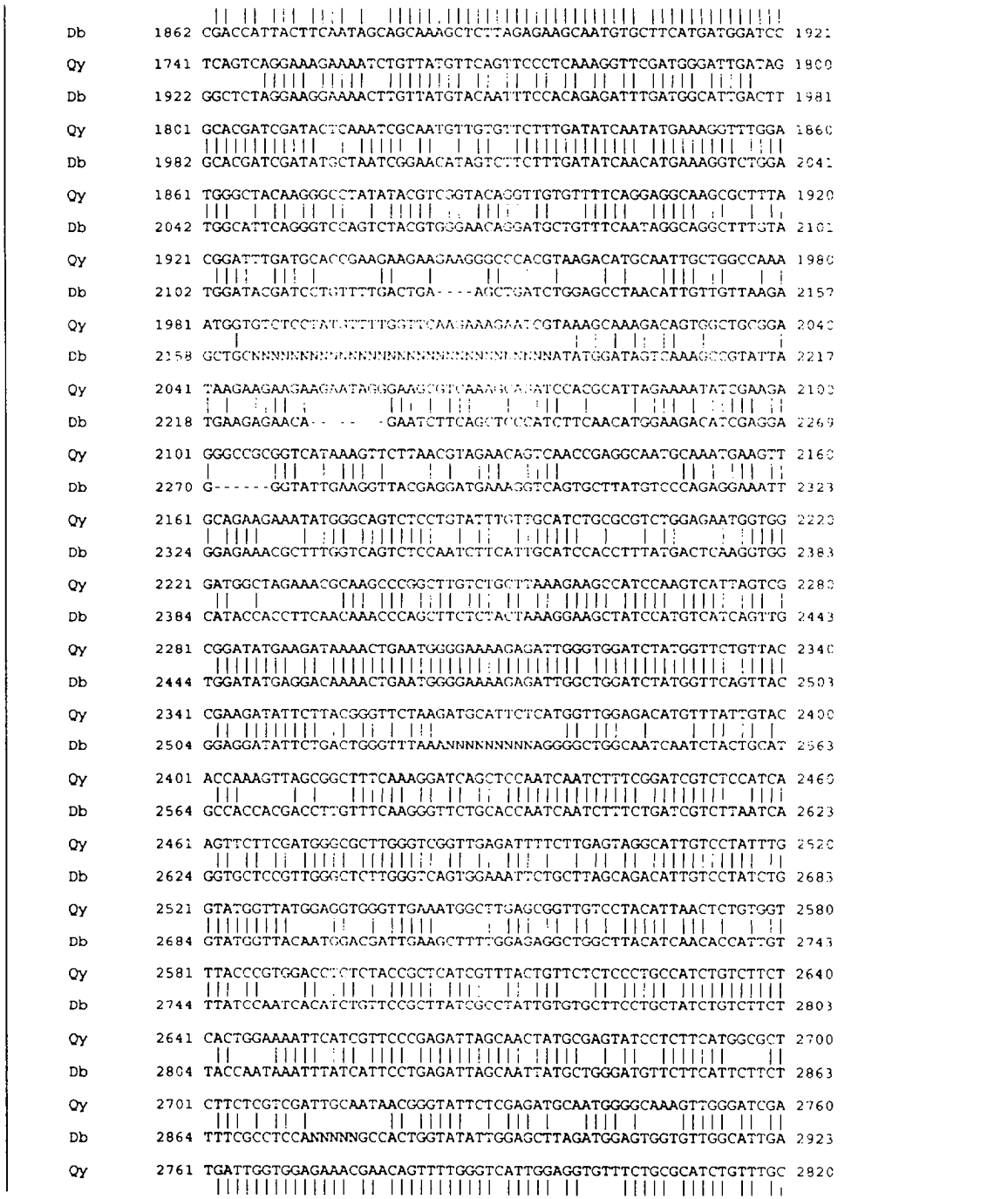
Qy         481 GATTTCTTCTGATAGACATGCTCTTATTGTTCCTCC[illegible]CATGGCAATAG 540

Db         665 AGAGATCCCTGATGCATCCCCTGACCGTCATTG[illegible]TCGAGCTATGT 724

Qy         541 AGTTCATCCTGTTTCTCTTTCTGACCCGACCGT[illegible]GATGGTACC 600

Db         725 TGATCCAAGCGTCCCAGTTCCTGTGAGGATTG[illegible]TGAATTCCTA 784
```





```
Db       2924 AGATTGGTGGAGAAATGAGCAGTTTTGGGTTATTGGTGGCACCTCTGCCCATCTCTTCGC 2983

Qy       2821 TCTCTTCCAAGGTCTCCTCAAGGTTCTTGCTGGTGTCGACACTAACTTCACAGTCACATC 2880

Db       2984 AGTGTTCCAGGGTCTGCTGAAAGTGTTGGCTGGGATTGATACCAACTTCACAGTTACCTC 3043

Qy       2881 AAAAGCAGCTGAT---GATGGAGAGTTCTCTGACCTTTACCTCTTCAAATGGACTTCACT 2937

Db       3044 AAAGGCATCTGATGAGGATGGCGACTTTGCTGAGCTATATGTGTTCAAGTGGACCAGTTT 3103

Qy       2938 TCTCATCCCTCCAATGACTCTACTCATCATAAACGTCATTGGAGTCATAGTCGGAGTCTC 2997

Db       3104 GCTCATTCCTCCGACCACTGTTCTTGTCATTAACCTGGTCGGAATGGTGGCAGGAATTTC 3163

Qy       2998 TGATGCCATCAGCAATGGATACGACTCGTGGGGACCGCTTTTCGGAAGACTGTTCTTTGC 3057

Db       3164 GTATGCCATTAACAGTGGCTACCAATCCTGGGGTCCGCTCTTTGGAAAGCTGTTCTTCTC 3223

Qy       3058 ACTTTGGGTCATCATTCATCTTTACCCGTTCCTTAAAGGTTTGCTTGGGAAACAAGATAG 3117

Db       3224 GATCTGGGTGNTCCTCCATCTGTACACGTTCCTCAAGGGTCTCATGGGAAGGCAGAACCG 3283

Qy       3118 AATGCCAACCATTATTGTCGTCTGGTCCATCCTCCTGGCCTCGATTCTTACACTTCTTTG 3177

Db       3284 CACACCAACAATCGTCATTGTCTGGTCCATCCTTCTTGCATCTATCTTCTCCTTGCTGTG 3343

Qy       3178 GGTCCGGGTTAATCCGTT 3195

Db       3344 GGTGAAGATCGATCCTTT 3361


RESULT 5
AY108113
LOCUS       AY108113               3763 bp    mRNA    linear   HTC 16-OCT-2002
DEFINITION  Zea mays PCO126465 mRNA sequence
ACCESSION   AY108113
VERSION     AY108113.1  GI:21211191
KEYWORDS    HTC.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 3763)
  AUTHORS   Rainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S.,
            Arthur,L.W., Hanafey,M., Krouse,M. and Tingey,S.V.
  TITLE     Maize Mapping Project (ZMP) of Consensus Sequences for Design of
            Overgo Probes
  JOURNAL   Unpublished (2002)
REFERENCE   2  (bases 1 to 3763)
  AUTHORS   Coe,E.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-APR-2002) Maize Mapping Project, University of
            Missouri, Columbia, MO 65211, USA
COMMENT     If you are interested in getting corresponding physical clones,
            these are publicly available from ZmDB and may be found by BLAST
            searching at MSL, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR,
            www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov  When the source of the
            maize cDNA sequences is either Virginia Walbot, Stanford or Pat
            Schnable, Iowa State, then clones may be requested from ZmDB:
            www.zmdb.iastate.edu.
FEATURES             Location/Qualifiers
     source          1. .3763
                     /organism="Zea mays"
                     /mol_type="mRNA"
                     /db_xref="MaizeDB:6373677"
                     /db_xref="taxon:4577"
                     /clone_lib="Maize Mapping Project/DuPont Consensus
                     Library"
                     /note="this sequence is part of a project of EST
                     assemblies resulting from the application of public
                     contigs to seed DuPont contigs; this resource was
                     assembled by DuPont as part of a collaboration for the
                     overgo addressing of BA[illegible] with the Maize
                     Mapping Project"
BASE COUNT      863 a    984 c   1011 g    9[illegible] c
ORIGIN

  Query Match             33.7%;  Score 1160;  [illegible] 763;
  Best Local Similarity   68.1%;  Pred. No. 2.
  Matches 1670;  Conservative    1;  Mismatch[illegible]  27;  Gaps    3;

Qy        748 AATGATGGATGAGGGAAGGCAGCCATTGTCTA[illegible] TCGAGCAAGAT 807

Db       1087 ATTGAACGACGAAACTCGACAGCCTCTATCTA[illegible] TCCTCCAGGAT 1146

Qy        808 AAATCCTTACCGGATGTTAATTGTGCTACGTC[illegible] TTCTTTCACTA 867

Db       1147 AAATCCATACAGGATGGTCATTGTGCTGCGA[illegible] TTCTTGCACTA 1206

Qy        868 CCGTATTCTTCACCCCGTCAAAGATGCATA[illegible] GTTATATGTGA 927

Db       1207 CCGTATCACAAATCCTGTGCGCAATGCATAC[illegible] GTTATATGTGA 1266

Qy        928 GATATGGTTTGCTGTTTCATGGGTTCTTGA[illegible] CCTATCGAGCG 987

Db       1267 GATCTGGTTTGCTCTTTCGTGGATATTGGA[illegible] CCAATCAACCG 1326

Qy        988 AGAAACGTACTTGGACCGACTCTCATTAAGA[illegible] CCGTCGGGACT 1047

Db       1327 GGAGACGTACCTTGATAGGCTGGCATTAAG[illegible] CCATCTCAGTT 1386

Qy       1048 ATCCCCTGTGGATGTATTTGTTAGTACAGTG[illegible] CCGCTTATTAC 1107

Db       1387 GGCTGCTGTTGACATTTTCGTCAGTACAGTC[illegible] CCTCTTGTCAC 1446

Qy       1108 TGCAAATACTGTCTTGTCTATTCTTGCTGTTG[illegible] GTTGCTTGTTA 1167

Db       1447 TGCCAATACCGTGCTATCCATTCTTGCTGTG[illegible] GTCTCTTGCTA 1506

Qy       1168 CGTATCTGATGATGGTGCTGCTATGCTTACT[illegible] ACCGCTGAATT 1227

Db       1507 TGTATCTGATGATGGAGCTGCGATGCTGACA[illegible] ACTTCAGAGTT 1566

Qy       1228 CGCAAGGAAATGGGTTCCTTTCTGCAAGAAA[illegible] GCTCCCGAATG 1287

Db       1567 TGCTAGAAAATGGGTACCATTTGTTAAGAAG[illegible] GCTCCTGAATG 1626

Qy       1288 GTATTTCTGCCATAAAATGGACTACTTGAA[illegible] TTGTTAGGGA 1347

Db       1627 GTACTTCTCCCAGAAAATTGATTACTTGAAG[illegible] TTGTTAAAGA 1686

Qy       1348 GCGGCGAGCCATGAAGAGAGATTATGAAGAA[illegible] GCTTTAGTAGC 1407

Db       1687 CCGCCGGGCCATGAAGAGAGAATATGAAGAA[illegible] GGCCTTGTTGC 1746

Qy       1408 AACAGCACAGAAAGTGCCTGAGGATGGTTGGA[illegible] CCTTGGCCCGG 1467

Db       1747 TAAGGCACAGAAAGTTCCTGAGGAAGGATGGA[illegible] CCATGGCCAGG 1806

Qy       1468 TAATAGTGTGCGAGATCATCCTGGCATGATTC[illegible] GACGGTGTTCG 1527

Db       1807 AAACAATACCMGGGACCATCCTGGAATGATTC[illegible] AGTGGTGGCCT 1866

Qy       1528 TGATGTCGAAAACAACGAGTTGCCTCGATTAG[illegible] AAGAGACCCGG 1587

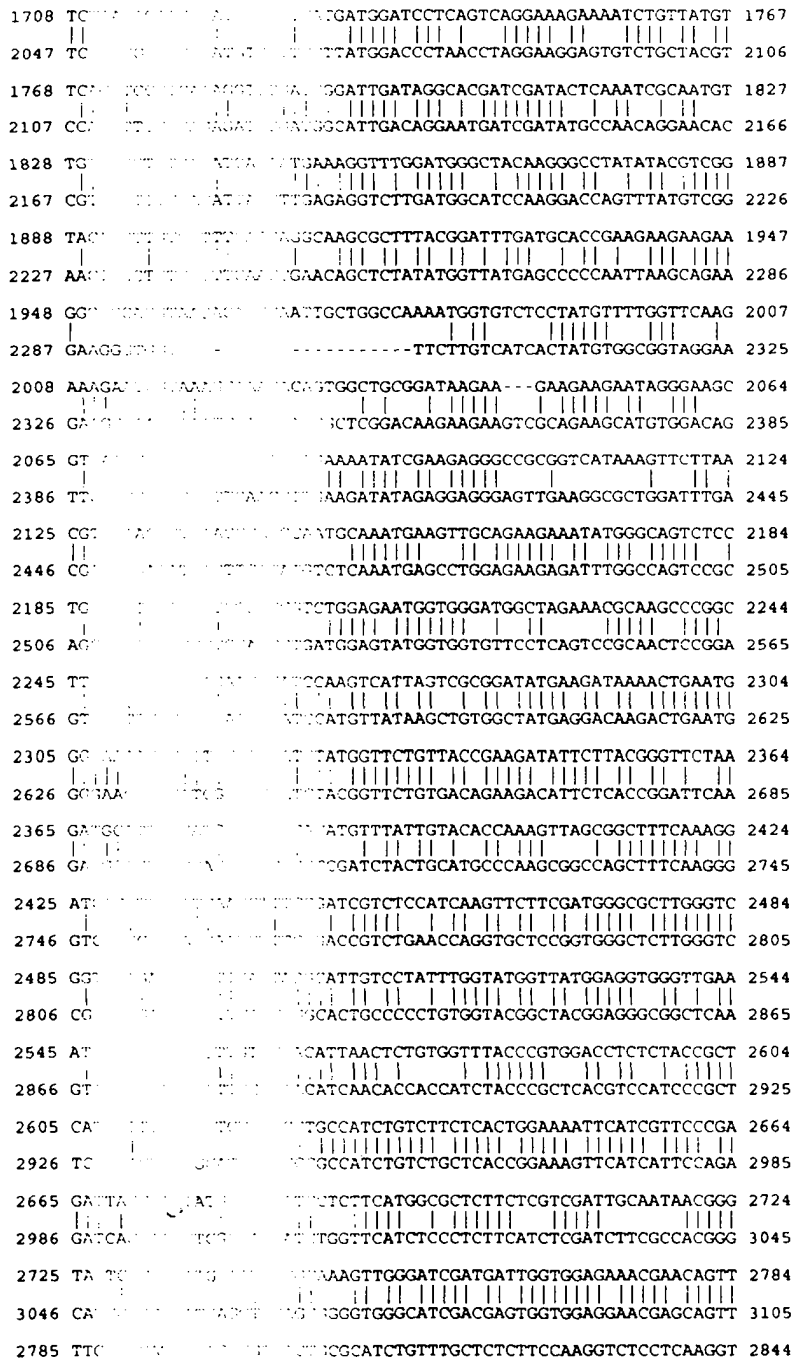
Db       1867 TGATACTGAGGGCAATGAGCTACCCCGTTTG[illegible] AAGCGTCCTGG 1926

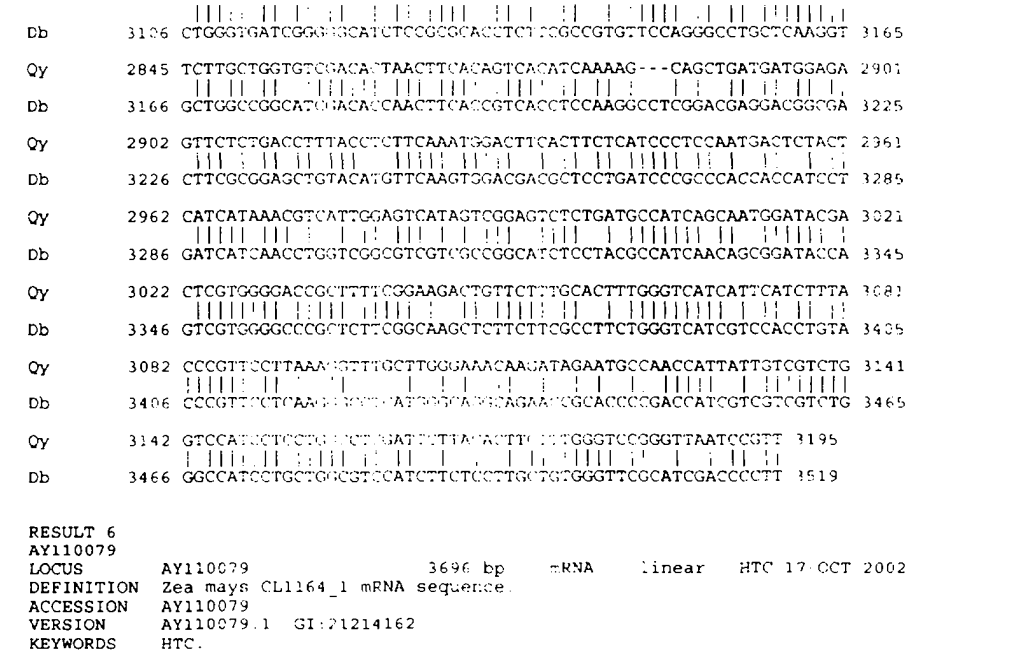
Qy       1588 ATTTGATCACCATAAGAAGGCTGGAGCTATGA[illegible] TCTGGGGTTCT 1647

Db       1927 ATTCCAGCATCACAAGAAAGCTGGTGCCATGA[illegible] TCAGCTGTGCT 1986

Qy       1648 ATCAAATGCTCCTTACCTTCTGAATGTCGACT[illegible] AATAGCAAAGC 1707

Db       1987 TACCAATGGACAATACATGTTGA[illegible] AACAGTAAGGC 2046
```

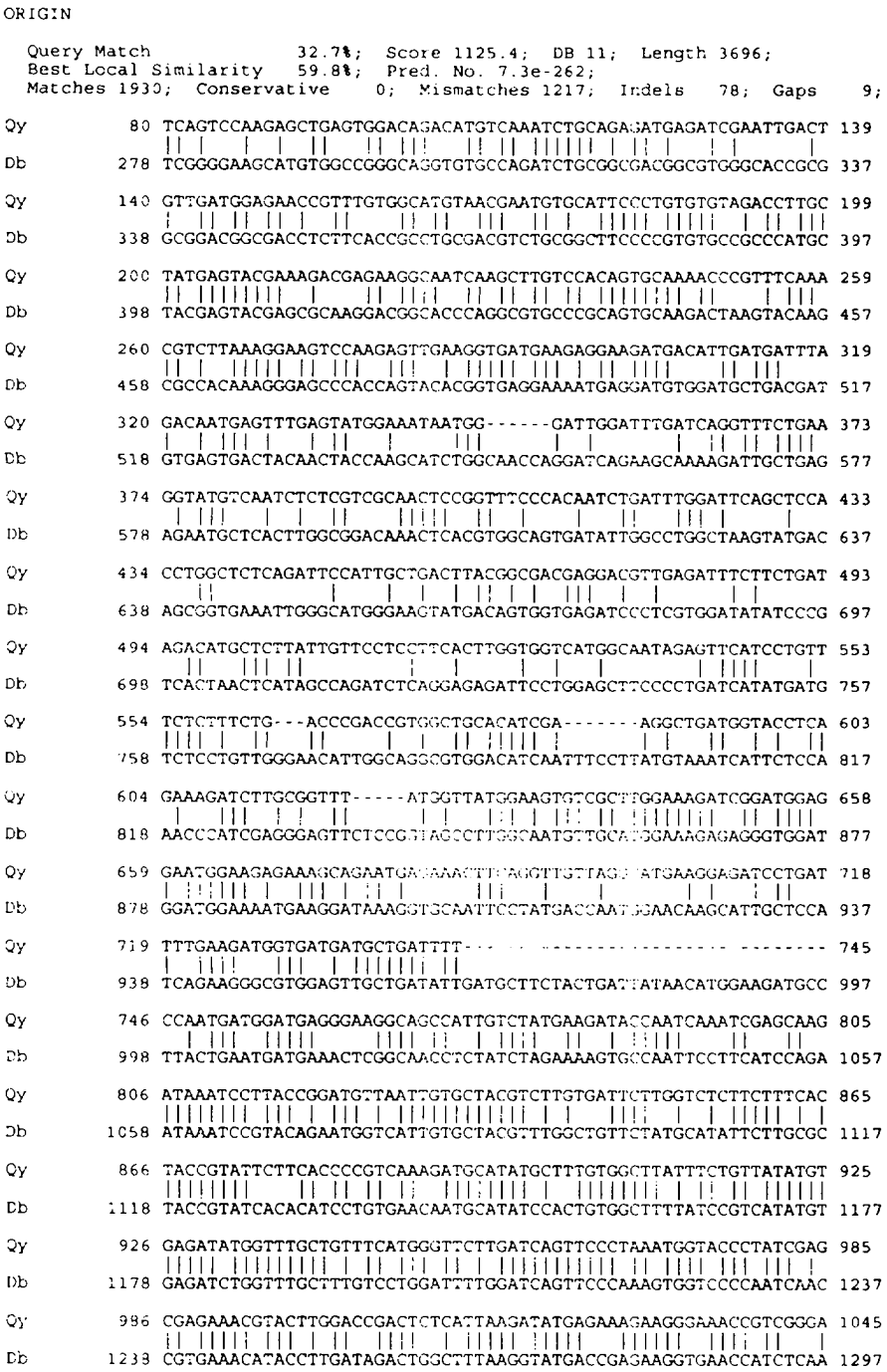


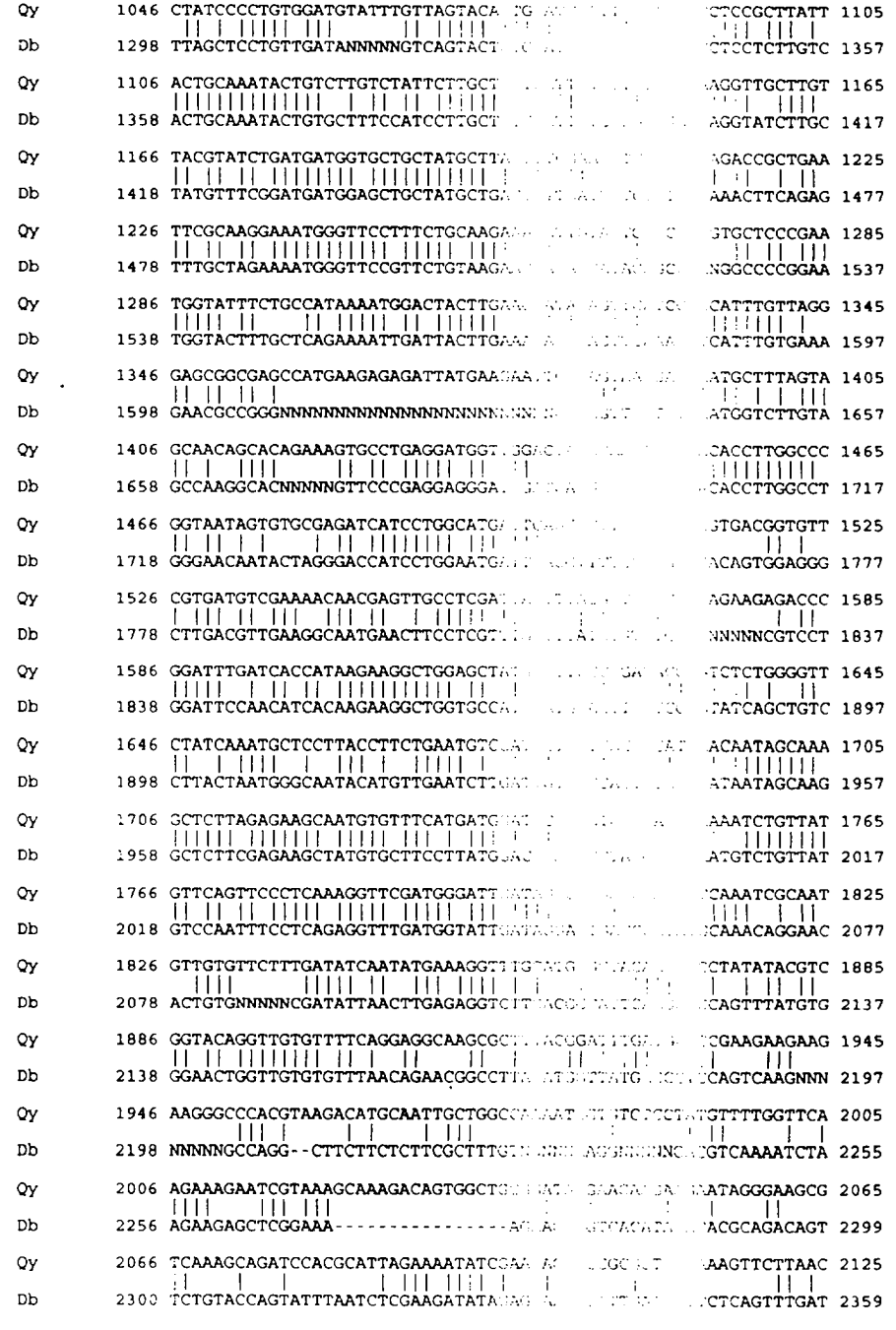


RESULT 6
AY110079
LOCUS AY110079 3696 bp mRNA linear HTC 17-OCT-2002
DEFINITION Zea mays CL1164_1 mRNA sequence.
ACCESSION AY110079
VERSION AY110079.1 GI:21214162
KEYWORDS HTC.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 3696)
AUTHORS Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S., Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
TITLE Maize Mapping Project/DuPont Consensus Sequences for Design of Overgo Probes
JOURNAL Unpublished (2002)
REFERENCE 2 (bases 1 to 3696)
AUTHORS Coe,E.H.
TITLE Direct Submission
JOURNAL Submitted (25-APR-2002) Maize Mapping Project, University of Missouri, Columbia, MO 65211, USA
COMMENT If you are interested in getting corresponding physical clones, these are publicly available from ZmDB and may be found by BLAST searching at MSL, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR, www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the maize cDNA sequences is either Virginia Walbot, Stanford or Pat Schnable, Iowa State, then clones may be requested from ZmDB: www.zmdb.iastate.edu.
FEATURES Location/Qualifiers
source 1..3696
/organism="Zea mays"
/mol_type="mRNA"
/db_xref="MaizeDB:630055"
/db_xref="taxon:4577"
/clone_lib="Maize Mapping Project/DuPont Cornsensus Library"
/note="this sequence is part of a project of EST assemblies resulting from the application of public contigs to seed DuPont contigs; this resource was assembled by DuPont as part of a collaboration for the overgo addressing of BACs in conjunction with the Maize Mapping Project"
BASE COUNT 831 a 766 c 880 g 945 t 274 others

ORIGIN

Query Match 32.7%; Score 1125.4; DB 11; Length 3696;
Best Local Similarity 59.8%; Pred. No. 7.3e-262;
Matches 1930; Conservative 0; Mismatches 1217; Indels 78; Gaps 9;





```
Qy       2126 GTAGA[illegible]ATGCAAATGAAGTTGCAGAAGAAATATGGGCAGTCTCCCT 2185
Db       2360 GATGA[illegible]TCTCAAATGAGCTTGGAGAAGAGATTTGGCCAGTCCAGT 2419

Qy       2186 GTATT[illegible]TGGAGAATGGTGGGATGGCTAGAAACGCAAGCCCGGCT 2245
Db       2420 GNNNNN[illegible]ATGGAATATGGTGGTGTTCCACAATCTGCAACTCCNNNN 2479

Qy       2246 TGTCTC[illegible]AAGTCATTAGTCGCGGATATGAAGATAAAACTGAATGG 2305
Db       2480 NNNNNN[illegible]ATGTCATCAGCTGTGGCTATGAGGACAAAACTGACTGG 2539

Qy       2306 GGAAA[illegible]GGTTCTGTTACCGAAGATATTCTTACGGGTTCTAAG 2365
Db       2540 GGAAC[illegible]ATGGTTCTGTTACAGAAGACATTCTCACCGGNNNNNNN 2599

Qy       2366 ATGCA[illegible]ATGTTTATTGTACACCAAAGTTAGCGGCTTTCAAAGGA 2425
Db       2600 NNNNN[illegible]NNNNNNNNNNNNNNCTAAGCGACCAGCTTTCAAGGGA 2659

Qy       2426 TCAGC[illegible]ATCGTCTCCATCAAGTTCTTCGATGGGCGCTTGGGTCG 2485
Db       2660 TCTGC[illegible]ATCGTTTGAATCAAGTGCTTCGGTGGGCTCTTGGTTCC 2719

Qy       2486 GTTGA[illegible]ATTGTCCTATTTGGTATGGTTATGGAGGTGGGTTGAAA 2545
Db       2720 ATTGA[illegible]ATTGTCCCATATGGTATGGCTATGGAGGCCGGCTTAAA 2779

Qy       2546 TGGCT[illegible]ATTAACTCTGTGGTTTACCCGTGGACCTCTCTACCGCTC 2605
Db       2780 TTCCT[illegible]TCAACACAACAATTTATCCACTCACATCAATCCCGCTC 2839

Qy       2606 ATCGT[illegible]CATCTGTCTTCTCACTGGAAAATTCATCGTTCCCGAG 2665
Db       2840 CTCCT[illegible]AGTTTGTCTTCTCACTGGGAAGTTCATCATCCCAAAG 2899

Qy       2666 ATTAG[illegible]TTCATGGCGCTCTTCTCGTCGATTGCAATAACGGGT 2725
Db       2900 ATTAG[illegible]GGTTTATATCGCTCTTTATCTCAATCTTTGCCACTGGT 2959

Qy       2726 ATTCT[illegible]AAGTTGGGATCGATGATTGGTGGAGAAACGAACAGTTT 2785
Db       2960 ATCCT[illegible]NNNNNNNCATTGATGAATGGTGGAGGAACGAGCAGTTC 3019

Qy       2786 TGGGT[illegible]GCATCTGTTTGCTCTCTTCCAAGGTCTCCTCAAGGTT 2845
Db       3020 TGGGT[illegible]CGCATTTATTTGCCGTCTTCCAGGGTCTCCTGAAGGTG 3079

Qy       2846 CTTGCT[illegible]TTCACAGTCACATCAAAAGCAGCT---GATGATGGAGAG 2902
Db       3080 CTTGCT[illegible]TTCACTGTCACCTCTAAGGCCACTGACGAAGAAGGTGAT 3139

Qy       2903 TTCTCT[illegible]AATGGACTTCACTTCTCATCCCTCCAATGACTCTACTC 2962
Db       3140 TTTGCC[illegible]AGTGGACAACGCTTCTGATCCCACCAACCACTATTTTG 3199

Qy       2963 ATCAT[illegible]AGTCGGAGTCTCTGATGCCATCAGCAATGGATACGAC 3022
Db       3200 ATCATC[illegible]GCTGGCATTTCCTACGCAATCAATAGCGGTTACCAG 3259

Qy       3023 TCGTG[illegible]ACTGTTCTTTGCACTTTGGGTCATCATTCATCTTTAC 3082
Db       3260 TCATG[illegible]AGCTCTTCTTTGCGTTCTGGGTGATTGTCCACCTGTAC 3319

Qy       3083 CCGTT[illegible]GGAAACAAGATAGAATGCCAACCATTATTGTCGTCTGG 3142
Db       3320 CCCTT[illegible]GGAAGCAGAACCGCACGCCGACCATTGTCGTTGTCTGG 3379

Qy       3143 TCCAT[illegible]TTACACTTCTTTGGGTCCGGGTTAATCCGTTTGTGGCG 3202
Db       3380 GCTAT[illegible]TTCCCTGATGTGGGTTCGTATCGATCCATTCACCACC 3439

Qy       3203 AAAGG[illegible]AGATCTGTGGTTTAGACTGCTTG 3244
Db       3440 CGGGTCACTGGTCCTGATATCGCGAAATGTGGCATCAACTGCTAG 3484
```

RESULT 7
AY104236
LOCUS AY104236 2872 bp mRNA linear HTC 16-OCT-2002
DEFINITION Zea mays PCO121439 mRNA sequence.
ACCESSION AY104236
VERSION AY104236.1 GI:21207314
KEYWORDS HTC.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 2872)
AUTHORS Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S., Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
TITLE Maize Mapping Project/DuPont Consensus Sequences for Design of Overgo Probes
JOURNAL Unpublished (2002)
REFERENCE 2 (bases 1 to 2872)
AUTHORS Coe,E.H.
TITLE Direct Submission
JOURNAL Submitted (25-APR-2002) Maize Mapping Project, University of Missouri, Columbia, MO 65211, USA
COMMENT If you are interested in getting corresponding physical clones, these are publicly available from ZmDB and may be found by BLAST searching at MSU, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR, www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the maize cDNA sequences is either Virginia Walbot, Stanford or Pat Schnable, Iowa State, then clones may be requested from ZmDB: www.zmdb.iastate.edu.
FEATURES Location/Qualifiers
source 1..2872
/organism="Zea mays"
/mol_type="mRNA"
/db_xref="MaizeDB:637064"
/db_xref="taxon:4577"
/clone_lib="Maize Mapping Project/DuPont Consensus Library"
/note="this sequence is part of a project of EST assemblies resulting from the application of public contigs to seed DuPont contigs; this resource was assembled by DuPont as part of a collaboration for the overgo addressing of BACs in conjunction with the Maize Mapping Project"
BASE COUNT 730 a 600 c 688 g 852 t 2 others
ORIGIN

Query Match 32.6%; Score 1123.8; DB 11; Length 2872;
Best Local Similarity 67.8%; Pred. No. 1.7e-261;
Matches 1653; Conservative 1; Mismatches 763; Indels 22; Gaps 5;

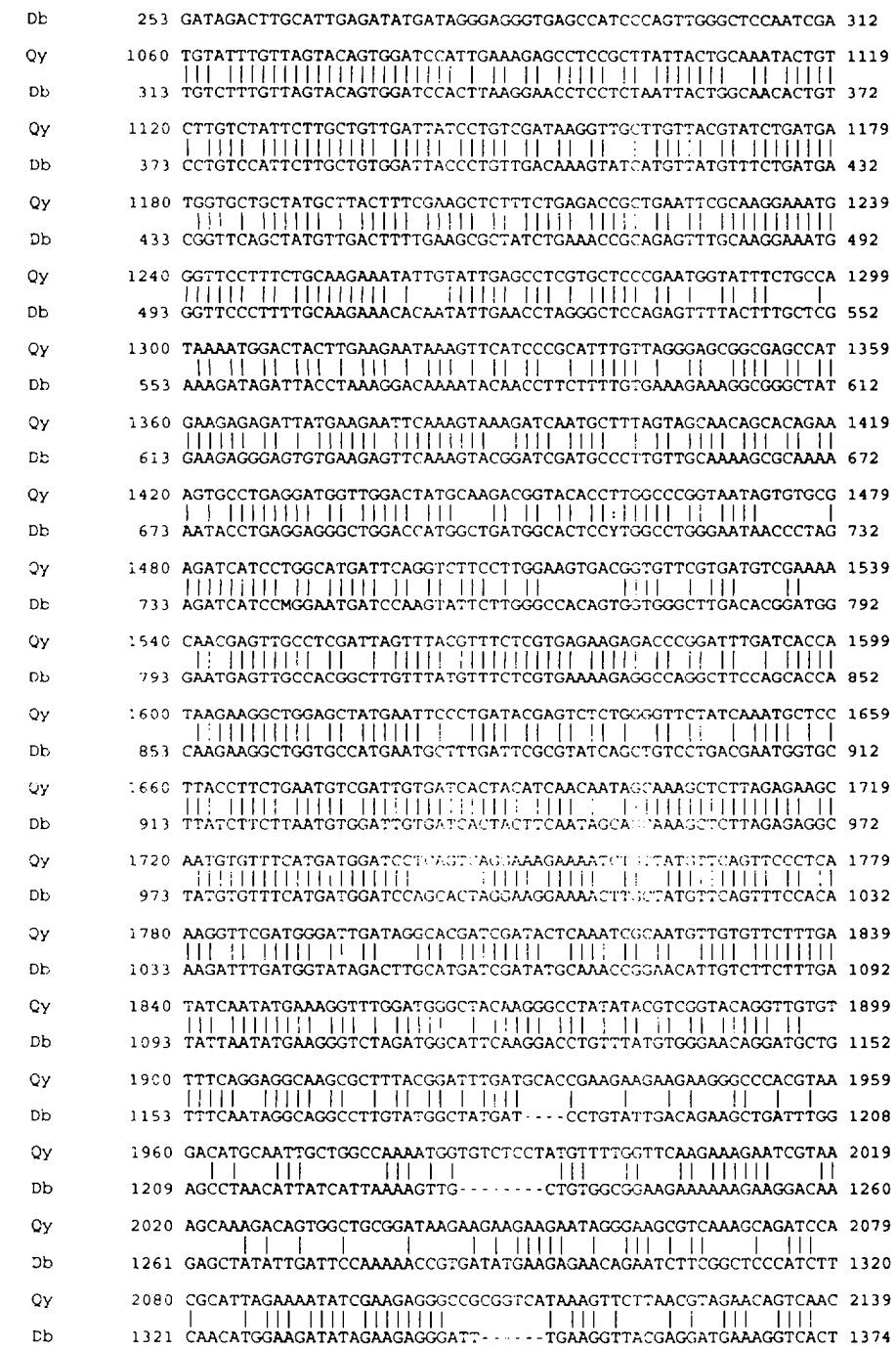
```
Qy        761 GGAAGGCAGCCATTGTCTATGAAGATACCAATCAAATCGAGCAAGATAAATCCTTACCGG 820
Db         13 GCACGGCTACCTCTAAGTCGCATAGTTCCGATATCTCCAAACGAGCTTAACCTTTATCGG 72

Qy        821 ATGTTAATTGTGCTACGTCTTGTGATTCTTGGTCTCTTCTTTCACTACCGTATTCTTCAC 880
Db         73 ATCGTGATTGTTCTCCGGCTTATCATCCTATGTTTCTTCTTTCAATATCGTATAACTCAT 132

Qy        881 CCCGTCAAAGATGCATATGCTTTGTGGCTTATTTCTGTTATATGTGAGATATGGTTTGCT 940
Db        133 CCAGTGGAAGATGCTTATGGGTTGTGGCTTGTATCTGTTATTTGTGAAGTTTGGTTTGCC 192

Qy        941 GTTTCATGGGTTCTTGATCAGTTCCCTAAATGGTACCCTATCGAGCGAGAAACGTACTTG 1000
Db        193 TTGTCTTGGCTTCTAGATCAGTTCCCAAAGTGGTATCCTATCAACCGTGAAACTTACCTC 252

Qy       1001 GACCGACTCTCATTAAGATATGAGAAAGAAGGGAAACCGT-CGGGACTATCCCCTGTGGA 1059
```


Db 253 GATAGACTTGCATTGAGATATGATAGGGAGGGTGAGCCATCCCAGTTGGGCTCCAATCGA 312
Qy 1060 TGTATTTGTTAGTACAGTGGATCCATTGAAAGAGCCTCCGCTTATTACTGCAAATACTGT 1119
Db 313 TGTCTTTGTTAGTACAGTGGATCCACTTAAGGAACCTCCTCTAATTACTGGCAACACTGT 372
Qy 1120 CTTGTCTATTCTTGCTGTTGATTATCCTGTCGATAAGGTTGCTTGTTACGTATCTGATGA 1179
Db 373 CCTGTCCATTCTTGCTGTGGATTACCCTGTTGACAAAGTATCATGTTATGTTTCTGATGA 432
Qy 1180 TGGTGCTGCTATGCTTACTTTCGAAGCTCTTTCTGAGACCGCTGAATTCGCAAGGAAATG 1239
Db 433 CGGTTCAGCTATGTTGACTTTTGAAGCGCTATCTGAAACCGCAGAGTTTGCAAGGAAATG 492
Qy 1240 GGTTCCTTTCTGCAAGAAATATTGTATTGAGCCTCGTGCTCCCGAATGGTATTTCTGCCA 1299
Db 493 GGTTCCCTTTTGCAAGAAACACAATATTGAACCTAGGGCTCCAGAGTTTTACTTTGCTCG 552
Qy 1300 TAAAATGGACTACTTGAAGAATAAAGTTCATCCCGCATTTGTTAGGGAGCGGCGAGCCAT 1359
Db 553 AAAGATAGATTACCTAAAGGACAAAATACAACCTTCTTTTGTGAAAGAAAGGCGGGCTAT 612
Qy 1360 GAAGAGAGATTATGAAGAATTCAAAGTAAAGATCAATGCTTTAGTAGCAACAGCACAGAA 1419
Db 613 GAAGAGGGAGTGTGAAGAGTTCAAAGTACGGATCGATGCCCTTGTTGCAAAAGCGCAAAA 672
Qy 1420 AGTGCCTGAGGATGGTTGGACTATGCAAGACGGTACACCTTGGCCCGGTAATAGTGTGCG 1479
Db 673 AATACCTGAGGAGGGCTGGACCATGGCTGATGGCACTCCYTGGCCTGGGAATAACCCTAG 732
Qy 1480 AGATCATCCTGGCATGATTCAGGTCTTCCTTGGAAGTGACGGTGTTCGTGATGTCGAAAA 1539
Db 733 AGATCATCCMGGAATGATCCAAGTATTCTTGGGCCACAGTGGTGGGCTTGACACGGATGG 792
Qy 1540 CAACGAGTTGCCTCGATTAGTTTACGTTTCTCGTGAGAAGAGACCCGGATTTGATCACCA 1599
Db 793 GAATGAGTTGCCACGGCTTGTTTATGTTTCTCGTGAAAAGAGGCCAGGCTTCCAGCACCA 852
Qy 1600 TAAGAAGGCTGGAGCTATGAATTCCCTGATACGAGTCTCTGGGGTTCTATCAAATGCTCC 1659
Db 853 CAAGAAGGCTGGTGCCATGAATGCTTTGATTCGCGTATCAGCTGTCCTGACGAATGGTGC 912
Qy 1660 TTACCTTCTGAATGTCGATTGTGATCACTACATCAACAATAGCAAAGCTCTTAGAGAAGC 1719
Db 913 TTATCTTCTTAATGTGGATTGTGATCACTACTTCAATAGCA AAAGCTCTTAGAGAGGC 972
Qy 1720 AATGTGTTTCATGATGGATCCTCAGTCAGAAAGAAAATCT TATGTTCAGTTCCCTCA 1779
Db 973 TATGTGTTTCATGATGGATCCAGCACTAGGAAGGAAAACTTGCTATGTTCAGTTTCCACA 1032
Qy 1780 AAGGTTCGATGGGATTGATAGGCACGATCGATACTCAAATCGCAATGTTGTGTTCTTTGA 1839
Db 1033 AAGATTTGATGGTATAGACTTGCATGATCGATATGCAAACCGGAACATTGTCTTCTTTGA 1092
Qy 1840 TATCAATATGAAAGGTTTGGATGGGCTACAAGGGCCTATATACGTCGGTACAGGTTGTGT 1899
Db 1093 TATTAATATGAAGGGTCTAGATGGCATTCAAGGACCTGTTTATGTGGGAACAGGATGCTG 1152
Qy 1900 TTTCAGGAGGCAAGCGCTTTACGGATTTGATGCACCGAAGAAGAAGAAGGGCCCACGTAA 1959
Db 1153 TTTCAATAGGCAGGCCTTGTATGGCTATGAT----CCTGTATTGACAGAAGCTGATTTGG 1208
Qy 1960 GACATGCAATTGCTGGCCAAAATGGTGTCTCCTATGTTTTGGTTCAAGAAAGAATCGTAA 2019
Db 1209 AGCCTAACATTATCATTAAAAGTTG--------CTGTGGCGGAAGAAAAAAGAAGGACAA 1260
Qy 2020 AGCAAAGACAGTGGCTGCGGATAAGAAGAAGAAGAATAGGGAAGCGTCAAAGCAGATCCA 2079
Db 1261 GAGCTATATTGATTCCAAAAACCGTGATATGAAGAGAACAGAATCTTCGGCTCCCATCTT 1320
Qy 2080 CGCATTAGAAAATATCGAAGAGGGCCGCGGTCATAAAGTTCTTAACGTAGAACAGTCAAC 2139
Db 1321 CAACATGGAAGATATAGAAGAGGGATT------TGAAGGTTACGAGGATGAAAGGTCACT 1374
Qy 2140 CGAGGCAATGCAAATGAAGTTGCAGAAGAA TTGTTGCATC 2199
Db 1375 GCTTATGTCTCAGAAGAGCTTGGAGAAA TTATTGCATC 1434
Qy 2200 TGCGCGTCTGGAGAATGGTGGGATGGCTA TGCTTAAAGA 2259
Db 1435 CACCTTTATGACTCAAGGTGGCATACCCC TGCTAAAGGA 1494
Qy 2260 AGCCATCCAAGTCATTAGTCGCGGATATG AAGAGATTGG 2319
Db 1495 AGCTATACATGTCATTAGTTGTGGATATG AAGAGATCGG 1554
Qy 2320 GTGGATCTATGGTTCTGTTACCGAAGATA ATTCTCATGG 2379
Db 1555 ATGGATATATGGCTCTGTTACTGAAGATA CATGCAAGAGG 1614
Qy 2380 TTGGAGACATGTTTATTGTACACCAAAGT GCTCCAATCAA 2439
Db 1615 TTGGATATCCATCTACTGCATGCCACTTC GCTCCGATTAA 1674
Qy 2440 TCTTTCGGATCGTCTCCATCAAGTTCTTCG AGATTTTCTT 2499
Db 1675 TCTTTCTGATCGTCTCAACCAAGTGTTACG AAATTCTACT 1734
Qy 2500 GAGTAGGCATTGTCCTATTTGGTATGGTTA CTTGAGCGGTT 2559
Db 1735 TAGCAGACACTGTCCTATCTGGTATGGTTA CTGGAGAGACT 1794
Qy 2560 GTCCTACATTAACTCTGTGGTTTACCCGT GTTTACTGTTC 2619
Db 1795 GGCATACATCAACACCATTGTTTATCCAA GCATACTGCGT 1854
Qy 2620 TCTCCCTGCCATCTGTCTTCTCACTGGAAA AGCAACTATGC 2679
Db 1855 CCTTCCTGCTATCTGTTTACTCACCAACAA AGCAATTATGC 1914
Qy 2680 GAGTATCCTCTTCATGGCGCTCTTCTCGTC CTCGAGATGCA 2739
Db 1915 TGGGGCGTTCTTCATCCTGCTTTTTGCTC TTGGAGCTTCG 1974
Qy 2740 ATGGGGCAAAGTTGGGATCGATGATTGGTC GTCATTGGAGG 2799
Db 1975 ATGGAGTGGTGTTGGCATTGAGGATTGGA GTCATTGGTGG 2034
Qy 2800 TGTTTCTGCGCATCTGTTTGCTCTCTTCG CTGGTGTCGA 2859
Db 2035 CACCTCTGCACATCTCTTTGCTGTGTTCG CAGGGATCGA 2094
Qy 2860 CACTAACTTCACAGTCACATCAAAAGCA- TCTGACCTTTA 2916
Db 2095 CACAAACTTCACGGTCACATCAAAGGCAACGA TGCTGAGCTGTA 2154
Qy 2917 CCTCTTCAAATGGACTTCACTTCTCATCCC ATAAACGTCAT 2976
Db 2155 TGTGTTCAAGTGGACAACTCTTCTGATCCC GATTAACCTGGT 2214
Qy 2977 TGGAGTCATAGTCGGAGTCTCTGATGCCAT TGGGGACCGCT 3036
Db 2215 TGGTATAGTGGCTGGAGTGTCGTATGCTAT ATGGGGTCCACT 2274
Qy 3037 TTTCGGAAGACTGTTCTTTGCACTTTGGGT TTCCTTAAAGG 3096
Db 2275 ATTCGGGAAGCTGTTCTTTGCAATCTGGGT TTCCTGAAGGG 2334
Qy 3097 TTTGCTTGGGAAACAAGATAGAATGCCAAC ATCCTCCTGGC 3156
Db 2335 TCTCATGGGGAAGCAGAACCGCACACCGAC GTCCTTCTTGC 2394
Qy 3157 CTCGATTCTTACACTTCTTTGGGTCCGGGT TT 3212
Db 2395 TTCCATATTCTCGCTGCTGTGGGTGAAGAT 2452

```
SULT 8
112236
CUS       AY11223:                    3728 bp    mRNA    linear   HTC 17-OCT-2002
FINITION  Zea mays                    quence.
CCESSION  AY11223(
RSION     AY11223(.
YWORDS    HTC.
URCE      Zea mays
ORGANISM  Zea mays
          Eukaryo                     Streptophyta; Embryophyta; Tracheophyta;
          Spermat                     yta; Liliopsida; Poales; Poaceae; PACCAD
          clade;                      opogoneae; Zea.
FERENCE   1  (bas
AUTHORS   Hainey,                     Mias,G.H., Vogel,J.M., Whitsitt,M.S.,
          Arthur,                     Morgante,M. and Tingey,S.V.
TITLE     Maize M                     Pont Consensus Sequences for Design of
          Overgo
JOURNAL   Unpubli
FERENCE   2  (bas
AUTHORS   Coe,E.H.
TITLE     Direct
JOURNAL   Submitt                     Maize Mapping Project, University of
          Missour                     65211, USA
MMENT     If you                      getting corresponding physical clones,
          these a                     le from ZmDB and may be found by BLAST
          searchi                     org; ZmDB, www.zmdb.iastate.edu; TIGR,
          www.tig                     .ncbi.nlm.nih.gov. When the source of the
          maize c                     ther Virginia Walbot, Stanford or Pat
          Schnabl                     n clones may be requested from ZmDB:
          www.zmd

ATURES                                ers
     source

                                      ys"

                                      DB:630049"

                                      4577"

                                      ze Mapping Project/DuPont Cornsensus

                                      uence is part of a project of EST
                                      lting from the application of public
                                      DuPont contigs; this resource was
                                      DuPont as part of a collaboration for the
                                      ing of BACs in conjunction with the Maize

SE COUNT      782                     92  g    878 t    298 others
IGIN

Query Match                           1102.4;  DB 11;  Length 3728;
Best Local Simila                     No. 2.8e-256;
Matches 1558;  Co                     Mismatches  872;  Indels   21;  Gaps    2;

,       748 AATGA                     CATTGTCTATGAAGATACCAATCAAATCGAGCAAGAT 807
            | | |                     | |||| |||  | ||| |    ||   || |||
>       993 ATTAA                     TTCTATCTAGGAAAGTTCCACTTCCTTCCTCCAGGAT 1052

,       808 AAATC                     GCTACGTCTTGTGATTCTTGGTCTCTTCTTTCACTA 867
            |||||  |                   |||||  |  |  ||||  |  |||||| |||||
>      1053 AAATC                     TGCTACGATTGATTGTTCTAAGCATCTTCTTGCACTA 1112

,       868 CCGTA                     AAGCATATGCTTTGTGGCTTATTTCTGTTATATGTGA 927
            |||  |                    |||||  |  |||||||| | ||||||||||||||
>      1113 CCGGA                     AAGCATACCCACTGTGGCTTCTATCTGTTATATGTGA 1172

,       928 GATAT                     TTCTTGATCAGTTCCCTAAATGGTACCCTATCGAGCG 987
            |||  |                    . |||||||| || || ||||  || |||  | ||
>      1173 GATCT                     TATTGGATCAGTTTCCAAAGTGGTTTCCAATCAACCG 1232

,       988 AGAAA                     TCTTATTAAGATATGAGAAAGAAGGGAAACCGTCGGGACT 1047
            || ||                     | | |||| |||||    |||||  | || ||    |
>      1233 CGAGA                     TTATTAAGGTATGACCGGGAAGGTGAGCCATCTCAGTT 1292

,      1048 ATCCC                     TTACAGTGGATCCATTGAAAGAGCCTCCGCTTATTAC 1107

Db     1293 GGCTGCTGTTGACANNNNNGTCAGTACTGTCGACCCAATGAAGGAGCCTCCTCTTGTCAC 1352

Qy     1108 TGCAAATACTGTCTTGTCTATTCTTGCTGTTGATTATCCTGTCGATAAGGTTGCTTGTTA 1167
            ||| |||||  ||    || |||||  |||| || |||||||| |||||||  |||| ||
Db     1353 TGCCAATACCGTGCTATCCATTCTCGCTGTGGACTATCCTGTGGATAAGGTCTCTTGCTA 1412

Qy     1168 CGTATCTGATGATGGTGCTGCTATGCTTACTTTCGAAGCTCTTTCTGAGACCGCTGAATT 1227
            ||||||||||||||  |||||||| | || || || ||  |||||||  | || ||
Db     1413 TGTATCTGATGATGGAGCTGCTATGCTGACATTTGATGCACTAGCTGAGACTTCAGAGTT 1472

Qy     1228 CGCAAGGAAATGGGTTCCTTTCTGCAAGAAATATTGTATTGAGCCTCGTGCTCCCGAATG 1287
            || || |||||| || ||    ||| |  |||| ||| | ||||| |||||
Db     1473 TGCTAGAAAATGGGTGCCATTTGTTAAGAAGTACAACATTGAACCTAGAGCTCCTGAATG 1532

Qy     1288 GTATTTCTGCCATAAAATGGACTACTTGAAGAATAAAGTTCATCCCGCATTTGTTAGGGA 1347
            ||| |||| |||  ||| || ||| |||| | |||| || ||  |||||||||  ||
Db     1533 GTACTTCTCCCAGAAAATTGATTACTTGAAGCACAAAGTGCACCCTTCATTTGTTAAAGA 1592

Qy     1348 GCGGCGAGCCATGAAGAGAGATTATGAAGAATTCAAAGTAAAGATCAATGCTTTAGTAGC 1407
            || || |                 |||| | | | |||   | || ||
Db     1593 CCGCCGGGNNNNNNNNNNNNNNNNNNNNNNNCAAAATTAGGGTAAATGGCCTTGTTGC 1652

Qy     1408 AACAGCACAGAAAGTGCCTGAGGATGGTTGGATTATGCAAGACGGTACACCTTGGCCCGG 1467
            |  ||||        || |||||  | | |  |||||||  || ||||| ||||| ||
Db     1653 TAAGGCACNNNNNNTCCCTGAGGAAGGATGGATCATGCAAGATGGCACACCATGGCCAGG 1712

Qy     1468 TAATAGTGTGCGAGATCATCCTGGCATGATTCAGGTCTTCCTTGGAAGTGACGGTGTTCG 1527
            ||  |  |  |.|||||||  |||||  |||| ||||||||          ||||  |
Db     1713 AAACAATACCAGGGACCATCCTGGAATGATTCAGGTTTTCCTTGGTCACAGTGGTGGTCT 1772

Qy     1528 TGATGTCGAAAACAACGAGTTGCCTCGATTAGTTTACGTTTCTCGTGAGAAGAGACCCGG 1587
            ||||  ||    || ||| | || || || || || ||||||||||       | || ||
Db     1773 TGATACTGAGGGTAATGAGCTACCCCGTTTGGTCTATGTTTCTCGTGNNNNNCGTCCTGG 1832

Qy     1588 ATTTGATCACCATAAGAAGGCTGGAGCTATGAATTCCCTGATACGAGTCTCTGGGGTTCT 1647
            ||| | || || ||||| ||||| || |||||| | ||  | || |||||  | || ||
Db     1833 ATTCCAGCATCACAAGAAAGCTGGTGCCATGAATGCTCTTGTCCGCGTCTCAGCTGTGCT 1892

Qy     1648 ATCAAATGCTCCTTACCTTCTGAATGTCGATTGTGATCACTACATCAACAATAGCAAAGC 1707
            | ||||  |  ||| |  |||| | |||||||||||||||||||||||||||| || ||
Db     1893 TACCAATGGACAATACATGTTGAATCTTGATTGTGATCACTACATCAACAACAGTAAGGC 1952

Qy     1708 TCTTAGAGAAGCAATGTGTTTCATGATGGATCCTCAGTCAGGAAAGAAAATCTGTTATGT 1767
            ||| || ||||| ||||| ||| | |||||||| |  |||||  ||  |||| ||||
Db     1953 TCTCAGGGAAGCTATGTGCTTCCTTATGGATCCTAACCTAGGAAGGAGTGTCTGCTATGT 2012

Qy     1768 TCAGTTCCCTCAAAGGTTCGATGGGATTGATAGGCACGATCGATACTCAAATCGCAATGT 1827
            |||||| || || |||||||||| |||||||  | |||||  | ||  | ||
Db     2013 TCAGTTTCCCCAGAGGTTCGATGGTATTGATAGGAATGATCGATATGCCAACAGGAACAC 2072

Qy     1828 TGTGTTCTTTGATATCAATATGAAAGGTTTGGATGGGCTACAAGGGCCTATATACGTCGG 1887
            |||      ||||| ||  ||| |||| | |||||  | |||||  || |  || || ||
Db     2073 CGTGNNNNNCGATATTAACTTGAGAGGTCTTGATGGCATCCAAGGACCAGTTTATGTGGG 2132

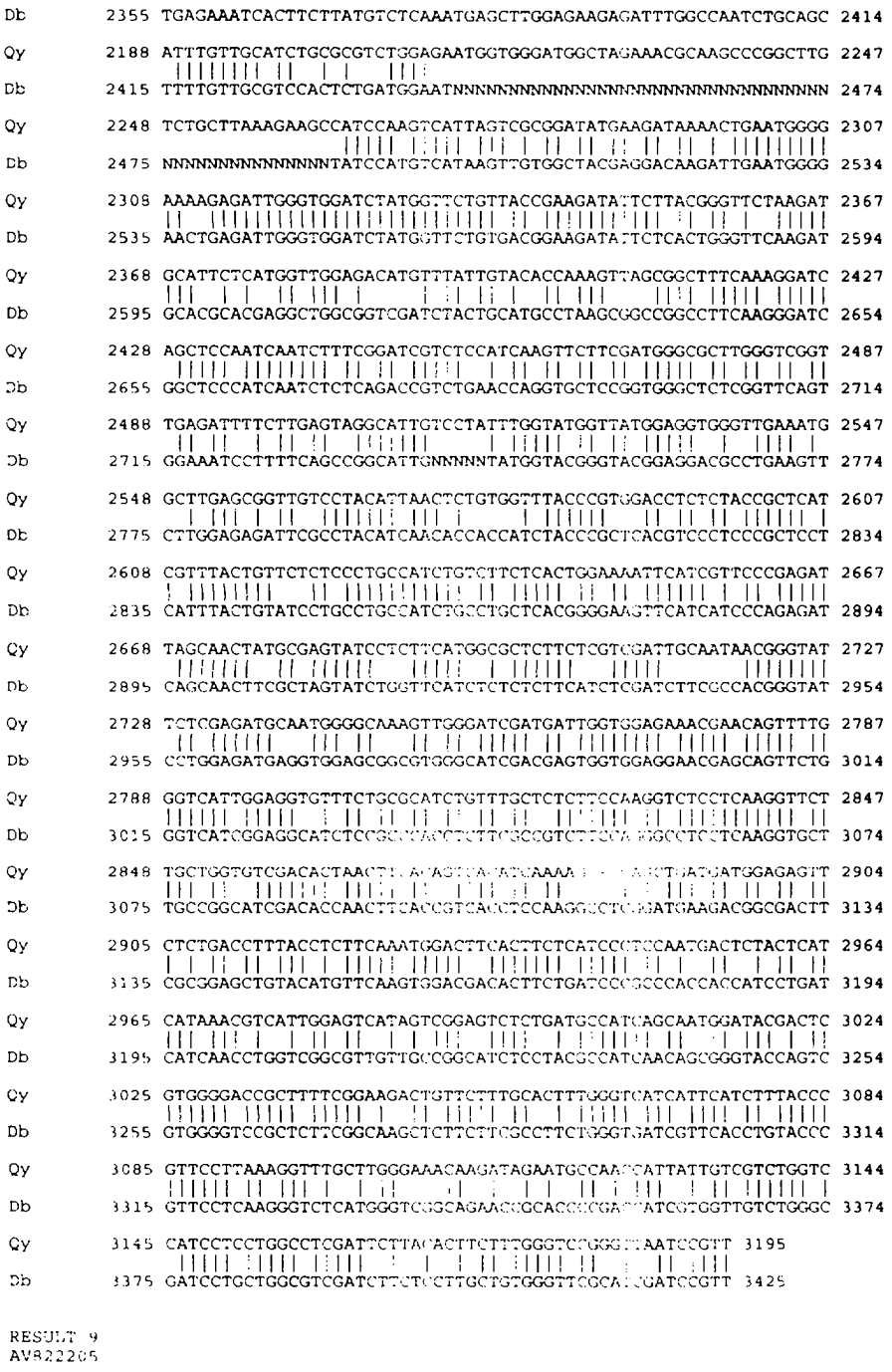
Qy     1888 TACAGGTTGTGTTTTCAGGAGGCAAGCGCTTTACGGATTTGATGCACCGAAGAAGAAGAA 1947
            || || ||||||||||  ||   ||| || ||  || | |||        |  ||| |  ||
Db     2133 CACTGGCTGTGTTTTCAACAGAACAGCTCTATATGGTTATGAGNNNNNAATTAAGCAAAA 2192

Qy     1948 GGGCCCACGTAAGACATGCAATTGCTGGCCAAAATGGTGTCTCCTATGTTTTGGTTCAAG 2007
            |                 ||  |  |  |  |  ||  |
Db     2193 G--------- --- -----AAGGGTGGTTTCTTGTCATCACTATGTGGNNNNNNNNNNNN 2234

Qy     2008 AAAGAATCGTAAAGCAAAGACAGTGGCTGCGGATAAGAAGAAGAAGAATAGGGAAGCGTC 2067
            |     |||     |  |  |
Db     2235 NNNNNNNNNNCAAAGAAGGGCTCAGACAAGNNNNNGNNNNNNNNNNNNNNNNNNNNN 2294

Qy     2068 AAAGCAGATCCACGCATTAGAAAATATCGAAGAGGGCCGCGGTCATAAAGTTCTTAACGT 2127
                                    | || ||||     |      |  |: | |
Db     2295 NNNNNNNNNNNNNNNNNNNNNNNNNTAGAGGAGGGAGTTGAAGGCGCTGGATTTGATGA 2354

Qy     2128 AGAACAGTCAACCGAGGCAATGCAAATGAAGTTGCAGAAGAAATATGGGCAGTCTCCTGT 2187
            ||  | |||              |||||||  ||| |||||| || ||| || ||| | |
```



RESULT 9
AV822205

```
LOCUS       AV822205                675 bp                    01-APR-2002
DEFINITION  AV822205 RAFL5 Arabidopsis thaliana                 02-P19 5',
            mRNA sequence.
ACCESSION   AV822205
VERSION     AV822205.1  GI:19864239
KEYWORDS    EST.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 675)
  AUTHORS   Seki,M., Narusaka,M., Ishida,J., Nanjo,T., Fujita,M., Oono,Y.,
            Kakajima,M., Oono,Y., Sakurai,T., Carninci,P., Hayashizaki,Y.,
            Arakawa,T., Shibata,K., Shinagawa,A., Yamamoto,M., Hayashizaki,Y.
            and Shinozaki,K.
  TITLE     Large scale analysis of Arabidopsis full-length cDNA (2002b)
  JOURNAL   Unpublished
COMMENT     Contact: Motoaki Seki
            Plant Functional Genomics Research Team
            RIKEN Genomic Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074 Japan
            Tel: 81-298-36-4359
            Fax: 81-298-36-9060
            Email: mseki@rtc.riken.go.jp
            An Arabidopsis full-length cDNA library was constructed essentially
            as reported previously (Seki et al., Plant J. 31:279-292 (2002)) in a
            modified pBluescript vector (pRAFL). Please visit our
            web site (http://www.gsc.riken.go.jp/Arabi/index.html) for
            further details.
FEATURES             Location/Qualifiers
     source          1..675
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
                     /clone="RAFL05-02-P19"
                     /dev_stage="rosette plant"
                     /lab_host="SOLR"
                     /clone_lib="RAFL5"
                     /note="Site_1: SstI; Site_2: NotI; Site_3: EcoRI;
                     dehydration-treated, 0.5 hour"
BASE COUNT      192 a    115 c    186 g    182 t      0 others
ORIGIN

  Query Match          18.5%;  Score 448;  DB 1;  Length 675;
  Best Local Similarity   97.3%;  Pred. No. 1.1e-39;
  Matches  658;  Conservative    0;  Mismatches   18;  Indels    1;  Gaps    1;

Qy       1900 TTTCAGGAGGCAAGCGCTTTACGGATTGATGCGAAAAGAGAAGGCACGTAA 1959
             ||||||||||||||||||||||| |||||| | ||| | |||||||
Db          1 TTTCAGGAGGCAAGCGCTTTAC-GATTGATCGAAAAGAGAAGGCCACGTAA 59

Qy       1960 GACATGCAATTGCTGGCCAAAATGGTGTCTCCTATTGGCTTTGATCGTAA 2019
             |||||||||||||||||||||||||||| ||| ||| | ||||||
Db         60 GACATGCAATTGCTGGCCAAAATGGTGTCTCCTATTGGCTTTGAATCGTAA 119

Qy       2020 AGCAAAGACAGTGGCTGCGGATAAGAAGAAGAAGGAAGCAAAGATCCA 2079
             |||||||||||||||||||||||||||||| ||| | ||||||||||
Db         120 AGCAAAGACAGTGGCTGCGGATAAGAAGAAGAAGGAAGCAAAGATCCA 179

Qy       2080 CGCATTAGAAAAATATCGAAGAGGGCCGCGGTTCTTCTTAAAAGTCAAC 2139
             |||||||||||||||||||||||||||| ||| | | | ||||||||||
Db         180 CGCATTAGAAAAATATCGAAGAGGGCCGCGTTCTTCTTAAAAGTCAAC 239

Qy       2140 CGAGGCAATGCAAATGAAGTTGCAGAAGAAGCAGCTCCTTTGTTGCATC 2199
             |||||||||||||||||||||| ||||| | |||| | |||||||
Db         240 CGAGGCAATGCAAATGAAGTTGGAGAAGAAGCAGCTCCTTTGTTGCATC 299

Qy       2200 TGCGCGTCTGGAGAATGGTGGGATGCTTAGAAAGTTGCTTTAAAGA 2259
             |||||| |||||||||||||||||||| | |||| | |||||||
Db         300 TGCGCGTATGGAGAATGGTGGGATGCTTAGAAAGTTGCTTTAAAGA 359

Qy       2260 AGCCATCCAAGTCATTAGTCGCGGATATGAAAAAAGAGATTGG 2319
```

```
Db        360 AGCCA  AT                 ATATGAAGATAAAACTGAATGGGGAAAAGAGATTGG 419

Qy       2320 GTGGA  AT                 GATATTCTTACGGGTTCTAAGATGCATTCTCATGG 2379

Db        420 GTGGA  AT                 GATATTCTTACGGGTTTTAAGATGCATTCTCATGG 479

Qy       2380 TTGGA  AT                 AAGTTAGCGGCTTTCAAAGGATCAGCTCCAATCAA 2439

Db        480 TTGGA  AT                 AAGTTAGCGGCTTTCGAAGGATCAGCTCCAATCAA 539

Qy       2440 TCTTT  AT                 TCTTCGATGGGCGCTTGGGTCGGTTGAGATTTTCTT 2499

Db        540 TCTTT  AT                 TCTTCGATGGGCGCTTGGGTCGGTTGAGATTTTCTT 599

Qy       2500 GAGTA  AT                 GTATGGTTATGGAGGTGGGTTGAAATGGCTTGAGCGGTT 2559

Db        600 GAGTA  AT                 TATGGTTATGGANGTGGGTTGAAATGGCTTGAGCGGNT 659

Qy       2560 GTCCT  AT
Db        660 GTCCT  AT

ESULT 10
V782896/c
OCUS       AV78289          bp    mRNA    linear   EST 28-MAR-2002
EFINITION  AV78289  thaliana cDNA clone RAFL05-02-P19 3',
           mRNA se
CCESSION   AV78289
ERSION     AV78289
EYWORDS    EST.
OURCE      Arabido  cress)
 ORGANISM  Arabido
           Eukary  reptophyta; Embryophyta; Tracheophyta;
           Sperma  ; eudicotyledons; core eudicots; rosids
           ; euros  Brassicaceae; Arabidopsis.
EFERENCE   1  (bas
 AUTHORS   Seki,M.  ka,J., Kamiya,A., Satou,M., Nakajima,M.,
           Oono,Y.  i,P., Kawai,J., Itoh,M., Ishii,Y.,
           Arakawa  nagawa,A., Muramatsu,M., Hayashizaki,Y.
           and Shi
 TITLE     Large s  bidopsis full-length cDNA (2002b)
 JOURNAL   Unpubl
OMMENT     Contact
           Plant F  esearch Group
           RIKEN G  er
           3-1-1 K  raki 305-0074, Japan
           Tel: 81
           Fax: 81
           Email:  p
           An Arab  cDNA library was constructed essentially
           as repo  et al., 1998). This clone is in a
           modifie  as a SstI/XhoI insert. Please visit our
           web sit  ken.go.jp/e/plant/index _e.html) for
           further
EATURES
     source
                    psis thaliana"
                  A"
                  2"
                  P19"
                  e plants"
                  R"
                  FL5"
                  SstI; Site_2: XhoI; subjected to
                  eated(1,2,5,10,24 hr)"
ASE COUNT      188          144 t
RIGIN

 Query Match          593;  DB 9;  Length 593;
 Best Local Simila    No. 8.6e-133;
 Matches  593;  Co    smatches    0;  Indels    0;  Gaps    0;

Qy       2797 AGGTGTTTCTGCGCATCTGTTTGCTCTCTTCCAAGGTCTCCTCAAGGTTCTTGCTGGTGT 2856

Db        593 AGGTGTTTCTGCGCATCTGTTTGCTCTCTTCCAAGGTCTCCTCAAGGTTCTTGCTGGTGT 534

Qy       2857 CGACACTAACTTCACAGTCACATCAAAAGCAGCTGATGATGGAGAGTTCTCTGACCTTTA 2916

Db        533 CGACACTAACTTCACAGTCACATCAAAAGCAGCTGATGATGGAGAGTTCTCTGACCTTTA 474

Qy       2917 CCTCTTCAAATGGACTTCACTTCTCATCCCTCCAATGACTCTACTCATCATAAACGTCAT 2976

Db        473 CCTCTTCAAATGGACTTCACTTCTCATCCCTCCAATGACTCTACTCATCATAAACGTCAT 414

Qy       2977 TGGAGTCATAGTCGGAGTCTCTGATGCCATCAGCAATGGATACGACTCGTGGGGACCGCT 3036

Db        413 TGGAGTCATAGTCGGAGTCTCTGATGCCATCAGCAATGGATACGACTCGTGGGGACCGCT 354

Qy       3037 TTTCGGAAGACTGTCTTTGCACTTTGGGTCATCATTCATCTTTACCCGTTCCTTAAAGG 3096

Db        353 TTTCGGAAGACTGTCTTTGCACTTTGGGTCATCATTCATCTTTACCCGTTCCTTAAAGG 294

Qy       3097 TTTGCTTGGGAAAAAGACAGAATGTAAACAATTATTGTCGTCTGGTCCATCCTCCTGGC 3156

Db        293 TTTGCTTGGGAAAAAGACAGAATGTAAACAATTATTGTCGTCTGGTCCATCCTCCTGGC 234

Qy       3157 CTCGATTCTTACACTTCTTTGGGTCCGGGTTAATCCGTTTGTGGCGAAAGGCGGTCCTAT 3216

Db        233 CTCGATTCTTACACTTCTTTGGGTCCGGGTTAATCCGTTTGTGGCGAAAGGCGGTCCTAT 174

Qy       3217 TCTCGAGATCTGTGGTTTAGACTGCTTGTGATTCGATTGACCGGTGGATGGGTTGGTGAA 3276

Db        173 TCTCGAGATCTGTGGTTTAGACTGCTTGTGATTCGATTGACCGGTGGATGGGTTGGTGAA 114

Qy       3277 AAAGGTTTAATTCCCACGGATCAAAGAGAGGTAAGAGAGATATTGTTTTACCTCTAAAAG 3336

Db        113 AAAGGTTTAATTCCCACGGATCAAAGAGAGGTAAGAGAGATATTGTTTTACCTCTAAAAG 54

Qy       3337 ACTCCTTCATTGTGTTCATTAGATGATGAAAAATGAAAAGAAAAAGAAGATTT 3389

Db         53 ACTCCTTCATTGTGTTCATTAGATGATGAAAAATGAAAAGAAAAAGAAGATTT 1

RESULT 11
AV823575
LOCUS       AV823575                582 bp    mRNA    linear   EST 01-APR-2002
DEFINITION  AV823575 RAFL5 Arabidopsis thaliana cDNA clone RAFL05-21-E11 5',
            mRNA sequence.
ACCESSION   AV823575
VERSION     AV823575.1  GI:19865635
KEYWORDS    EST.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 582)
  AUTHORS   Seki,M., Narusaka,M., Ishida,J., Kamiya,A., Satou,M., Nakajima,M.,
            Oono,Y., Sakurai,T., Carninci,P., Kawai,J., Itoh,M., Ishii,Y.,
            Arakawa,T., Shibata,K., Shinagawa,A., Muramatsu,M., Hayashizaki,Y.
            and Shinozaki,K.
  TITLE     Large scale analysis of Arabidopsis full-length cDNA (2002b)
  JOURNAL   Unpublished
COMMENT     Contact: Motoaki Seki
            Plant Functional Genomics Research Group
            RIKEN Genomic Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: 81-298-36-4359
            Fax: 81-298-36-9060
            Email: mseki@rtc.riken.go.jp
            An Arabidopsis full-length cDNA library was constructed essentially
            as reported previously (Seki et al., 1998).This clone is in a
            modified pBluescript vector as a SstI/XhoI insert. Please visit our
            web site (http://www.gsc.riken.go.jp/e/plant/index _e.html) for
```

```
          further details.
FEATURES             Location/Qualifiers
     source          1. .582
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
                     /clone="RAFL05-21-E11"
                     /dev_stage="rosette plants"
                     /lab_host="SOLR"
                     /clone_lib="RAFL5"
                     /note="Site_1: SstI; Site_2: XhoI; subjected to
                     dehydration-treated(1,2,5,10,24 hr)"
BASE COUNT      151 a    114 c    143 g    171 t      3 others
ORIGIN

  Query Match             15.8%;  Score 543;  DB 9;  Length 582;
  Best Local Similarity   99.5%;  Pred. No. 1.2e-120;
  Matches  565;  Conservative    0;  Mismatches    1;  Indels    2;  Gaps    2;

Qy       1026 AAGAAGGGAAACCGTCGGGACTATCCCCTGTGGATGTATTTGTTAGTACAGTGGATCCAT 1085
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db             1 AAGAAGGGAAACCGTCGGGACTATCCCCTGTGGATGTATTTGTTAGTACAGTGGATCCAT 60

Qy       1086 TGAAAGAGCCTCCGCTTATTACTGCAAATACT-GTCTTGTCTATTCTTGCTGTTGATTAT 1144
             |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db            61 TGAAAGAGCCTCCGCTTATTACTGCAAATACTAGTCTTGTCTATTCTTGCTGTTGATTAT 120

Qy       1145 CCTGTCGATAAGGTTGCTTGTTACGTATCTGATGATGGTGCTGCTATGCTTACTTTCGAA 1204
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCTGTCGATAAGGTTGCTTGTTACGTATCTGATGATGGTGCTGCTATGCTTACTTTCGAA 180

Qy       1205 GCTCTTTCTGAGACCGCTGAATTCGCAAGGAAATGGGTTCCTTTCTGCAAGAAATATTGT 1264
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GCTCTTTCTGAGACCGCTGAATTCGCAAGGAAATGGGTTCCTTTCTGCAAGAAATATTGT 240

Qy       1265 ATTGAGCCTCGTGCTCCCGAATGGTATTTCTGCCATAAAATGGACTACTTGAAGAATAAA 1324
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ATTGAGCCTCGTGCTCCCGAATGGTATTTCTGCCATAAAATGGACTACTTGAAGAATAAA 300

Qy       1325 GTTCATCCCGCATTTGTTAGGGAGCGGCGAGCCATGAAGAGAGATTATGAAGAATTCAAA 1384
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GTTCATCCCGCATTTGTTAGGGAGCGGCGAGCCATGAAGAGAGATTATGAAGAATTCAAA 360

Qy       1385 GTAAAGATCAATGCTTTAGTAGCAACAGCACAGAAAGTGCTTAGGATGGTTGGACTATG 1444
             |||||||||||| |||| |||| | |||||||||||||||||||||||||||||||||||||
Db        361 GTAAAGATCAATGCTTTAGTAGCAACAGCACAGAAAGTGCTTAGGATGGTTGGACTATG 420

Qy       1445 CAAGACGGTACACCTTGGCCCGGTAATAGTGTGCGAGATCATCCTGGCATGATTCAGGTC 1504
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CAAGACGGTACACCTTGGCCCGGTAATAGTGTGCGAGATCATCCTGGCATGATTCAGGTC 480

Qy       1505 TTCCTTGGAAGTGACGGTGTTCGTGATGT-CGAAAACAACGAGTTGCCTCGATTAGTTTA 1563
             |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db        481 TTCCTTGGAAGTGACGGTGTTCGTGATGTCCGAAAACAACGAGTTGCCTCGATTAGTTTA 540

Qy       1564 CGTTTCTCGTGAGAAGAGACCCGGATTT 1591
             ||||||||||||||||||||||||||||
Db        541 CGTTTCTCGTGAGAAGAGACCCGGATNT 568


RESULT 12
AV524115/c
LOCUS       AV524115                637 bp    mRNA    linear   EST 01-SEP-2000
DEFINITION  AV524115 Arabidopsis thaliana aboveground organs two to six-week
            old Arabidopsis thaliana cDNA clone APZL53c11F 3', mRNA sequence.
ACCESSION   AV524115
VERSION     AV524115.1  GI:8683643
KEYWORDS    EST.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 637)
  AUTHORS   Asamizu,E., Nakamura,Y., Sato [illegible] ata,S.
  TITLE     A large scale analysis of cDN[illegible] Generation
            of 12,028 non-redundant expressed s[illegible] tags fro[illegible] ormalized and
            size-selected cDNA libraries
  JOURNAL   DNA Res. 7, 175-180 (2000)
  MEDLINE   20363093
   PUBMED   10907847
COMMENT     Contact: Erika Asamizu
            The First Laboratory for Plan[illegible] Gen[illegible] search
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba [illegible] an
            Email: asamizu@kazusa.or.jp, [illegible] ;p/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .637
                     /organism="Arabidopsis tha[illegible]
                     /mol_type="mRNA"
                     /strain="Columbia"
                     /db_xref="taxon:370[illegible]"
                     /clone="APZL53c11F"
                     /tissue_type="above[illegible]"
                     /dev_stage="two to six-[illegible]"
                     /clone_lib="Arabido[illegible] above[illegible] rgans two to
                     six-week old"
                     /note="Vector: pBlue[illegible]; Site_2:
                     XhoI"
BASE COUNT      221 a    140 c    105 g    171 [illegible]
ORIGIN

  Query Match             14.5%;  Score [illegible];  [illegible] Length 637;
  Best Local Similarity   100.0%;  Pred. [illegible]
  Matches  500;  Conservative    0;  Mis[illegible]  In[illegible]  Gaps    0;

Qy       2937 TTCTCATCCCTCCAATGACTCTACTCA[illegible]TGGA[illegible]GGAGTCT 2996
             |||||||||||||||||||||||||||[illegible]||||||
Db        637 TTCTCATCCCTCCAATGACTCTACTCATCATA[illegible]TGGA[illegible]GGAGTCT 578

Qy       2997 CTGATGCCATCAGCAATGGATACGAC[illegible]GGTGGA[illegible]TTTC[illegible]GTTCTTTG 3056
             |||||||||||||||||||||||||||[illegible]||||||
Db        577 CTGATGCCATCAGCAATGGATACGAC[illegible]GGTGA[illegible]TTTC[illegible]GTTCTTTG 518

Qy       3057 CACTTTGGGTCATCATTCATCTTTAC[illegible]GGTTCCT[illegible]TTTG[illegible]AAAGATA 3116
             |||||||||||||||||||||||||||[illegible]||||||
Db        517 CACTTTGGGTCATCATTCATCTTTACCCGTT[illegible]TTTG[illegible]AAAGATA 458

Qy       3117 GAATGCCAACCATTATTGTCGTCTGGT[illegible]CCTC[illegible]ACTTCTTT 3176
             |||||||||||||||||||||||||||[illegible]||||||
Db        457 GAATGCCAACCATTATTGTCGTCTGGT[illegible]CCTC[illegible]ACTTCTTT 398

Qy       3177 GGGTCCGGGTTAATCCGTTTGTGGCGAAAG[illegible]TTCTC[illegible]GGTTTAG 3236
             |||||||||||||||||||||||||||[illegible]||||||
Db        397 GGGTCCGGGTTAATCCGTTTGTGGCGAAAGGCG[illegible]TTCTCGAGATC[illegible]GGTTTAG 338

Qy       3237 ACTGCTTGTGATTCGATTGACCGGTG[illegible]GGT[illegible]AAAAGGTTTAA[illegible]CCACGGA 3296
             |||||||||||||||||||||||||||[illegible]||||||
Db        337 ACTGCTTGTGATTCGATTGACCGGTGGATGGGT[illegible]AAAAGGTTTAA[illegible]CCACGGA 278

Qy       3297 TCAAAGAGAGGTAAGAGAGATATTGTTTTACCTC[illegible]AGACTCCTT[illegible]TGTTCATT 3356
             |||||||||||||||||||||||||||[illegible]||||||
Db        277 TCAAAGAGAGGTAAGAGAGATATTGTTTTACCT[illegible]AGACTCCTT[illegible]TGTTCATT 218

Qy       3357 AGATGATGAAAAATGAAAAGAAAAAGAAGATTTAA[illegible]TGTTACGAGAATT[illegible]TATTTTT 3416
             |||||||||||||||||||||||||||[illegible]||||||
Db        217 AGATGATGAAAAATGAAAAGAAAAAGAAGATTTAA[illegible]TGTTACGAGAATT[illegible]TATTTTT 158

Qy       3417 GCAAGAATGTGTTGTAGATA 3436
             ||||||||||||||||||||
Db        157 GCAAGAATGTGTTGTAGATA 138


RESULT 13
AV527222
LOCUS       AV527222                510 bp    [illegible]  lin[illegible]  [illegible]T 01-SEP-2000
```

```
EFINITION  AV5272[illegible] thaliana aboveground organs two to six-week
           old Ara[illegible] ana cDNA clone APZ31h04R 5', mRNA sequence.
CCESSION   AV5272[illegible]
ERSION     AV5272.[illegible]
EYWORDS    EST.
OURCE      Arabido[illegible] (thale cress)
 ORGANISM  Arabido[illegible]
           Eukaryo[illegible] Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
           Spermat[illegible] Magnoliophyta; eudicotyledons; core eudicots; rosids
           ; euros[illegible] Capparales; Brassicaceae; Arabidopsis.
EFERENCE   1  (bas[illegible]
 AUTHORS   Asamizu[illegible] E.,[illegible] Nakamura,Y.,[illegible] Sato,S. and Tabata,S.
 TITLE     A larg[illegible] cDNA in Arabidopsis thaliana: Generation
           of 12,[illegible] expressed sequence tags from normalized and
           size-s[illegible] libraries
 JOURNAL   DNA Re[illegible] 13 (2), 101-106 (2000)
 MEDLINE   203630[illegible]
  PUBMED   109078[illegible]
OMMENT     Contac[illegible] Kazusa DNA Research
           The Fir[illegible] Plant Gene Research
           Kazusa[illegible] ute
           Yana 1[illegible] a 292-0812, Japan
           Email:[illegible] URL:http://www.kazusa.or.jp/en/plant/.
EATURES                    Location/Qualifiers
     source                1..510
                           /organism="Arabidopsis thaliana"
                           /mol_type="mRNA"
                           /cultivar="cv. Columbia"
                           /db_xref="taxon:3702"
                           /clone="APZ31h04R"
                           /tissue_type="aboveground organs"
                           /dev_stage="six-week old"
                           /clone_lib="Arabidopsis thaliana aboveground organs two to
                           six-week old"
                           /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
ASE COUNT      1[illegible]                148 t
RIGIN

  Query Match             1.2%;  Score 487.6;  DB 9;  Length 510;
  Best Local Similarity   [illegible]  Pred. No. 3.3e-107;
  Matches  496;  [illegible]  Mismatches   14;  Indels    0;  Gaps    0;

Qy       2066 TCAA[illegible] ATATCGAAGAGGGCCGCGGTCATAAAGTTCTTAAC 2125
              ||||           ||||||||||||||||    ||||| | ||||
Db          1 TCAA[illegible] ATATCGAAGAGGGCCGCGTCACTAAAGGTTCTAAC 60

Qy       2126 GTAGA[illegible] CAAATGAAGTTGCAGAAGAAATATGGGCAGTCTCCT 2185
              ||||           ||||||||| |||||||| ||||||||||||||||
Db         61 GTAG[illegible] CAAATGAAGTTGGAGAAGAAATTTGGGCAGTCTCCT 120

Qy       2186 GTAT[illegible] GAGAATGGTGGGATGGCTAGAAACGCAAGCCCGGCT 2245
              ||||           ||||||||||||||||||||||||||||||||||||
Db        121 GTAT[illegible] AGAATGGTGGGATGGCTAGAAACGCAAGCCCGGCT 180

Qy       2246 TGTC[illegible] TCATTAGTCGCGGATATGAAGATAAAACTGAATGG 2305
              ||||           ||||||||| ||||||||||||||||||||||||
Db        181 TGTC[illegible] TCATTAGTTGCGGATATGAAGATAAAACTGAATGG 240

Qy       2306 GGAA[illegible] GTTCTGTTACCGAAGATATTCTTACGGGTTCTAAG 2365
              |||            |||||||||||||||||||||||||||| ||||
Db        241 GGAA[illegible] GGTTCTGTTACCGAAGATATTCTTACGGGTTTTAAG 300

Qy       2366 AT[illegible] GTTCATTGTACACCAAAGTTAGCGGCTTTCAAAGGA 2425
              |||            |||||||||||||||||||||||||||||||||||
Db        301 AT[illegible] GTTCATTGTACACCAAAGTTAGCGGCTTTCAAAGGA 360

Qy       2426 T[illegible] TCCATCAAGTTCTTCGATGGGCGCTTGGGTCG 2485
              |||            |||||||||||||||||||||||||||||||||
Db        361 TC[illegible] CGTTCCATCAAGTTCTTCGATGGGCGCTTGGGTCG 420

Qy       2486 GT[illegible] TGTCCTATTTGGTATGGTTATGGAGGTGGGTTGAAA 2545
              ||||           |||||||||||||||||||||||||||||||||||

Db         421 GTTGAGATTTTCTTGAGTAGGCATTGTCCTATTTGGTATGGTTATGGAGGTGGGTTGAAA 480

Qy       2546 TGGCTTGAGCGGTTGTCCTACATTAACTCT 2575
              |||||||||||||||||||||||||||||
Db         481 TGGCTTGAGCGGTTGTCCTACATTAACTCT 510

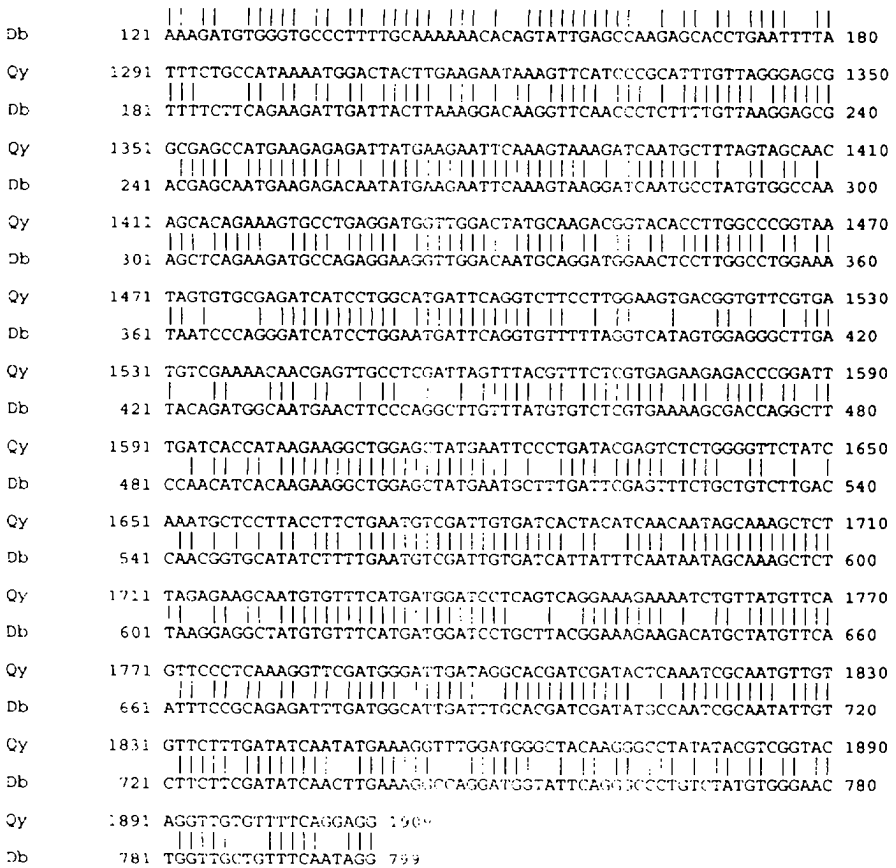

RESULT 14
CB894714
LOCUS       CB894714                 799 bp    mRNA    linear   EST 24-APR-2003
DEFINITION  EST647506 HOGA Medicago truncatula cDNA clone HOGA-32K17, mRNA
            sequence.
ACCESSION   CB894714
VERSION     CB894714.1  GI:30101883
KEYWORDS    EST.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 799)
  AUTHORS   Hahn,M.G., Cyamus Reuhs,T., Damaso,D., Town,C.D., Van Aken,S.,
            Utterback,T., Tho,J. and Fraser,C.M.
  TITLE     ESTs from roots of Medicago truncatula treated with
            oligogalacturonides of DP 6-20
  JOURNAL   Unpublished
COMMENT     Contact: Michael G. Hahn
            Complex Carbohydrate Research Center
            University of Georgia
            220 Riverbend Road, Athens, GA 30602-4712, USA
            Tel: 706-542-4457
            Fax: 706-542-4412
            Email: hahn@ccrc.uga.edu
            TIGR sequence name: MTMDW69TK
            More information is available at: www.medicago.org
            Seq primer: SKmod (CTA gAA CTA gtg gAT CC).
FEATURES              Location/Qualifiers
     source           1..799
                      /organism="Medicago truncatula"
                      /mol_type="mRNA"
                      /cultivar="A17"
                      /db_xref="taxon:3880"
                      /clone="HOGA-32K17"
                      /tissue_type="3 day old seedling roots"
                      /dev_stage="24 hours after treatment in the dark at 26 C
                      with 0.5 mg/ml oligogalacturonides (DP 6-20) in the
                      presence of 100 ug/ml Gentamicin"
                      /lab_host="XLOLR"
                      /clone_lib="HOGA"
                      /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                      XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA
                      was directionally ligated into the Unizap XR vector from
                      Stratagene and packaged using Gigapack III Gold packaging
                      extracts. Plasmids containing cDNA inserts were excised
                      from the recombinant lambda-Zap phage using Ex-assist
                      helper phage and propagated in SOLR cells."
BASE COUNT      220 a    143 c    193 g    243 t
ORIGIN

  Query Match             14.0%;  Score 483.8;  DB 14;  Length 799;
  Best Local Similarity   75.3%;  Pred. No. 3.1e-106;
  Matches  602;  Conservative    0;  Mismatches  197;  Indels    0;  Gaps    0;

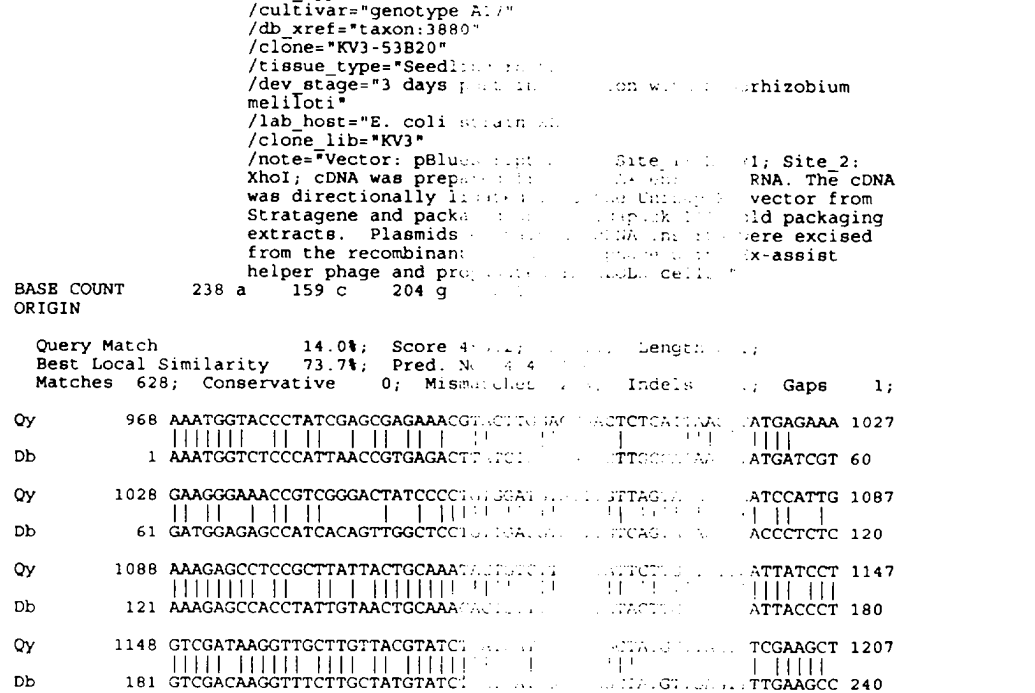
Qy       1111 AAATACTGTCTTGTCTATTCTTGCTGTTGATTATCCTGTCGATAAGGTTGCTTGTTACGT 1170
              ||| |||||| |||||||| ||||| ||||||| :|||||||| ||:||| |||| || ||
Db          1 AAACACTGTTTTGTCTATACTTGCCGTTGATTTTCCTGTCGACAAGGTTTCTTGCTATGT 60

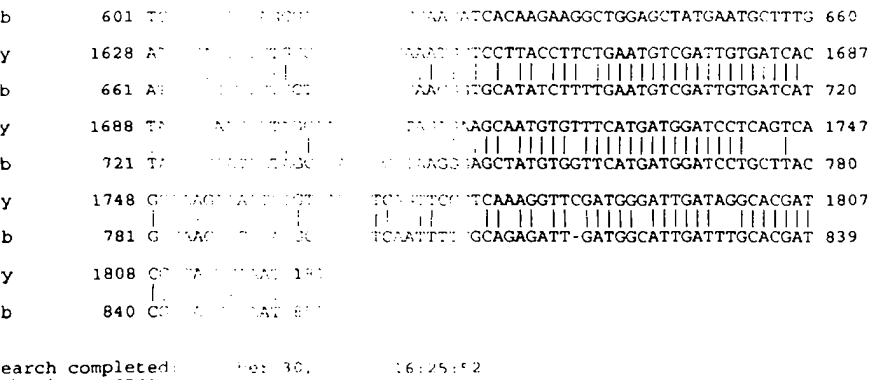
Qy       1171 ATCTGATGATGGTGCTGCTATGCTTACTTTCGAAGCTCTTTCTGAGACCGCTGAATTCGC 1230
              |||||||||||||| | |||||| | | ||| |||||  | ||||| || || || || ||
Db         61 ATCTGATGATGGTTCAGCTATGTTGAGTTTTGAAGCCTTATCTGAAACAGCCGAGTTTGC 120

Qy       1231 AAGGAAATGGGTTCCTTTCTGCAAGAAATATTGTATTGAGCCTCGTGCTCCCGAATGGTA 1290
```



RESULT 15
CB892280
LOCUS CB892280 851 bp mRNA linear EST 24-APR-2003
DEFINITION EST649249 KV3 Medicago truncatula cDNA clone KV3-53B20, mRNA sequence.
ACCESSION CB892280
VERSION CB892280.1 GI:30099448
KEYWORDS EST.
SOURCE Medicago truncatula (barrel medic)
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids ; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.
REFERENCE 1 (bases 1 to 851)
AUTHORS VandenBosch,K., Endre,G., Hur,J., Beremand,P., Town,C.D., Van Aken ,S., Utterback,T., Cho,J. and Fraser,C.M.
TITLE ESTs from roots of Medicago truncatula 72 h after Rhizobium inoculation, 2001
JOURNAL Unpublished
COMMENT Contact: VandenBosch K
Department of Plant Biology
University of Minnesota
220 BioSci Center, 1445 Gortner Ave, St. Paul, MN 55108, USA
Tel: 612 624 2755
Fax: 612 625 1738
Email: kvandenb@cbs.umn.edu
TIGR sequence name: MTEDF10TK
More information is available [illegible] ...ayo.org
Seq primer: SKmod (CTA gAA CTA [illegible]
FEATURES Location/Qualifiers
source 1. .851
/organism="Medicago t[illegible]
/mol_type="mRNA"
/cultivar="genotype A[illegible]"
/db_xref="taxon:3880"
/clone="KV3-53B20"
/tissue_type="Seedl[illegible]
/dev_stage="3 days [illegible] ...on w[illegible] ...rhizobium meliloti"
/lab_host="E. coli [illegible]
/clone_lib="KV3"
/note="Vector: pBlu[illegible] Site_[illegible] 1; Site_2: XhoI; cDNA was prep[illegible] RNA. The cDNA was directionally l[illegible] vector from Stratagene and packa[illegible] ld packaging extracts. Plasmids [illegible] ere excised from the recombinan[illegible] x-assist helper phage and pro[illegible] cells."
BASE COUNT 238 a 159 c 204 g [illegible]
ORIGIN





earch completed: [illegible] 30, 16:25:52
ob time : 6769.[illegible]

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 30, 2003, 12:47:22 ; Search time 865.127 Seconds
(without alignments)
10827.272 Million cell updates/sec

Title: US-09-980-043-5
Perfect score: 3444
Sequence: 1 gcggccgcggttaatcgccg.....gtgttgtagatagcggccgc 3444

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1811591 seqs, 1359896290 residues

Total number of hits satisfying chosen parameters: 3623182

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3211.2	93.2	3255	10	US-09-938-842A-2530	Sequence 2530, Ap
2	1944	56.4	3255	10	US-09-938-842A-559	Sequence 559, App
3	1934.4	56.2	3828	14	US-10-229-193-7	Sequence 7, Appli
4	1489.8	43.3	3936	10	US-09-900-237-9	Sequence 9, Appli
5	1489.8	43.3	3969	12	US-10-209-059-49	Sequence 49, Appl
6	1489.8	43.3	3969	12	US-10-160-719-17	Sequence 17, Appl
7	1489.8	43.3	3969	12	US-10-160-719-37	Sequence 37, Appl
8	1489	43.2	3776	10	US-09-900-237-3	Sequence 3, Appli
9	1464.8	42.5	3568	12	US-10-209-059-41	Sequence 41, Appl
10	1464.8	42.5	3568	12	US-10-160-719-1	Sequence 1, Appli
11	1458.4	42.3	2890	10	US-09-900-237-17	Sequence 17, Appl
12	1416.4	41.1	3786	10	US-09-900-237-7	Sequence 7, Appli
13	1416.4	41.1	3813	12	US-10-209-059-17	Sequence 17, Appl
14	1416.4	41.1	3813	12	US-10-160-719-25	Sequence 25, Appl
15	1416.4	41.1	3813	12	US-10-160-719-45	Sequence 45, Appl
16	1347	39.1	2663	10	US-09-900-237-23	Sequence 23, Appl
17	1294.8	37.6	3603	14	US-10-229-193-5	Sequence 5, Appli
18	1294.8	37.6	3673	14	US-10-229-193-11	Sequence 11, Appl
19	1252.6	36.4	3614	14	US-10-229-193-9	Sequence 9, Appli
20	1234	35.8	3563	10	US-09-900-237-25	Sequence 25, Appl
21	1212.8	35.2	3725	12	US-10-209-059-45	Sequence 45, Appl
22	1212.8	35.2	3725	12	US-10-160-719-13	Sequence 13, Appl
23	1212.8	35.2	3725	12	US-10-160-719-21	Sequence 21, Appl
24	1212.8	35.2	3725	12	US-10-160-719-41	Sequence 41, Appl
25	1206.2	35.0	3753	12	US-10-160-719-33	Sequence 33, Appl
26	1206.2	35.0	3753	12	US-10-160-719-53	Sequence 53, Appl
27	1206.2	35.0	3780	12	US-10-209-059-1	Sequence 1, Appli
28	1206.2	35.0	3780	12	US-10-160-719-9	Sequence 9, Appli
29	1205.8	35.0	3517	10	US-09-900-237-13	Sequence 13, Appl
30	1205.4	35.0	3704	12	US-10-209-059-13	Sequence 13, Appl
31	1205.4	35.0	3704	12	US-10-160-719-57	Sequence 57, Appl
32	1177	34.2	3626	10	US-09-900-237-29	Sequence 29, Appl
33	1168.4	33.9	3799	12	US-10-267-459-5	Sequence 5, Appli
34	1168.4	33.9	3799	12	US-10-209-059-21	Sequence 21, Appl
35	1158.8	33.6	3746	12	US-10-160-719-29	Sequence 29, Appl
36	1158.8	33.6	3746	12	US-10-160-719-49	Sequence 49, Appl
37	1158.8	33.6	3773	12	US-10-209-059-9	Sequence 9, Appli
38	1158.8	33.6	3773	12	US-10-160-719-5	Sequence 5, Appli
39	1145	33.2	3328	10	US-09-838-539-1	Sequence 1, Appli
40	1137.4	33.0	2830	12	US-10-267-459-1	Sequence 1, Appli
41	1137.4	33.0	2830	12	US-10-209-059-5	Sequence 5, Appli
42	1111.6	32.3	3231	12	US-10-209-059-27	Sequence 27, Appl
43	1020	29.6	3028	12	US-10-209-059-29	Sequence 29, Appl
44	943.2	27.4	3470	12	US-10-209-059-25	Sequence 25, Appl
45	822.2	23.9	2248	14	US-10-229-193-1	Sequence 1, Appli

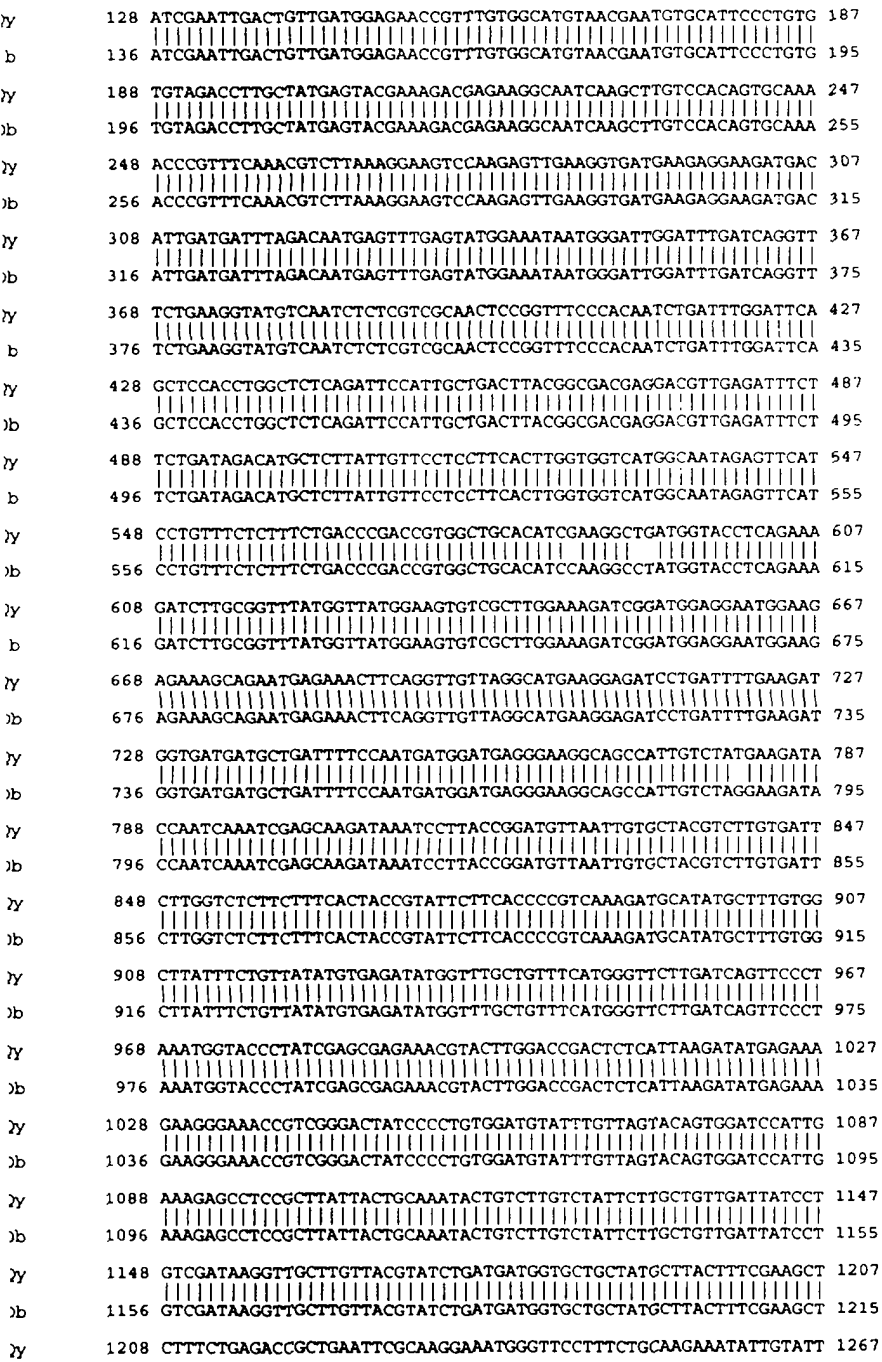
ALIGNMENTS

RESULT 1
US-09-938-842A-2530
; Sequence 2530, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRIP1300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 2530
; LENGTH: 3255
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-2530

Query Match 93.2%; Score 3211.2; DB 10; Length 3255;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 3222; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

```
Qy          8 CGGTTAATCGCCGGTTCTCACAACAGGAATGAGTTTGTCCTCATTAATGCCGATGAGAAT 67
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         16 CGGTTAATCGCCGGTTCTCACAACAGGAATGAGTTTGTCCTCATTAATGCCGATGAGAAT 75

Qy         68 GCCCGAATAAGATCAGTCCAAGAGCTGAGTGGACAGACATGTCAAATCTGCAGAGATGAG 127
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         76 GCCCGAATAAGATCAGTCCAAGAGCTGAGTGGACAGACATGTCAAATCTGCAGAGATGAG 135
```


Qy 128 ATCGAATTGACTGTTGATGGAGAACCGTTTGTGGCATGTAACGAATGTGCATTCCCTGTG 187
Db 136 ATCGAATTGACTGTTGATGGAGAACCGTTTGTGGCATGTAACGAATGTGCATTCCCTGTG 195
Qy 188 TGTAGACCTTGCTATGAGTACGAAAGACGAGAAGGCAATCAAGCTTGTCCACAGTGCAAA 247
Db 196 TGTAGACCTTGCTATGAGTACGAAAGACGAGAAGGCAATCAAGCTTGTCCACAGTGCAAA 255
Qy 248 ACCCGTTTCAAACGTCTTAAAGGAAGTCCAAGAGTTGAAGGTGATGAAGAGGAAGATGAC 307
Db 256 ACCCGTTTCAAACGTCTTAAAGGAAGTCCAAGAGTTGAAGGTGATGAAGAGGAAGATGAC 315
Qy 308 ATTGATGATTTAGACAATGAGTTTGAGTATGGAAATAATGGGATTGGATTTGATCAGGTT 367
Db 316 ATTGATGATTTAGACAATGAGTTTGAGTATGGAAATAATGGGATTGGATTTGATCAGGTT 375
Qy 368 TCTGAAGGTATGTCAATCTCTCGTCGCAACTCCGGTTTCCCACAATCTGATTTGGATTCA 427
Db 376 TCTGAAGGTATGTCAATCTCTCGTCGCAACTCCGGTTTCCCACAATCTGATTTGGATTCA 435
Qy 428 GCTCCACCTGGCTCTCAGATTCCATTGCTGACTTACGGCGACGAGGACGTTGAGATTTCT 487
Db 436 GCTCCACCTGGCTCTCAGATTCCATTGCTGACTTACGGCGACGAGGACGTTGAGATTTCT 495
Qy 488 TCTGATAGACATGCTCTTATTGTTCCTCCTTCACTTGGTGGTCATGGCAATAGAGTTCAT 547
Db 496 TCTGATAGACATGCTCTTATTGTTCCTCCTTCACTTGGTGGTCATGGCAATAGAGTTCAT 555
Qy 548 CCTGTTTCTCTTTCTGACCCGACCGTGGCTGCACATCGAAGGCTGATGGTACCTCAGAAA 607
Db 556 CCTGTTTCTCTTTCTGACCCGACCGTGGCTGCACATCCAAGGCCTATGGTACCTCAGAAA 615
Qy 608 GATCTTGCGGTTTATGGTTATGGAAGTGTCGCTTGGAAAGATCGGATGGAGGAATGGAAG 667
Db 616 GATCTTGCGGTTTATGGTTATGGAAGTGTCGCTTGGAAAGATCGGATGGAGGAATGGAAG 675
Qy 668 AGAAAGCAGAATGAGAAACTTCAGGTTGTTAGGCATGAAGGAGATCCTGATTTTGAAGAT 727
Db 676 AGAAAGCAGAATGAGAAACTTCAGGTTGTTAGGCATGAAGGAGATCCTGATTTTGAAGAT 735
Qy 728 GGTGATGATGCTGATTTTCCAATGATGGATGAGGGAAGGCAGCCATTGTCTATGAAGATA 787
Db 736 GGTGATGATGCTGATTTTCCAATGATGGATGAGGGAAGGCAGCCATTGTCTAGGAAGATA 795
Qy 788 CCAATCAAATCGAGCAAGATAAATCCTTACCGGATGTTAATTGTGCTACGTCTTGTGATT 847
Db 796 CCAATCAAATCGAGCAAGATAAATCCTTACCGGATGTTAATTGTGCTACGTCTTGTGATT 855
Qy 848 CTTGGTCTCTTCTTTCACTACCGTATTCTTCACCCCGTCAAAGATGCATATGCTTTGTGG 907
Db 856 CTTGGTCTCTTCTTTCACTACCGTATTCTTCACCCCGTCAAAGATGCATATGCTTTGTGG 915
Qy 908 CTTATTTCTGTTATATGTGAGATATGGTTTGCTGTTTCATGGGTTCTTGATCAGTTCCCT 967
Db 916 CTTATTTCTGTTATATGTGAGATATGGTTTGCTGTTTCATGGGTTCTTGATCAGTTCCCT 975
Qy 968 AAATGGTACCCTATCGAGCGAGAAACGTACTTGGACCGACTCTCATTAAGATATGAGAAA 1027
Db 976 AAATGGTACCCTATCGAGCGAGAAACGTACTTGGACCGACTCTCATTAAGATATGAGAAA 1035
Qy 1028 GAAGGGAAACCGTCGGGACTATCCCCTGTGGATGTATTTGTTAGTACAGTGGATCCATTG 1087
Db 1036 GAAGGGAAACCGTCGGGACTATCCCCTGTGGATGTATTTGTTAGTACAGTGGATCCATTG 1095
Qy 1088 AAAGAGCCTCCGCTTATTACTGCAAATACTGTCTTGTCTATTCTTGCTGTTGATTATCCT 1147
Db 1096 AAAGAGCCTCCGCTTATTACTGCAAATACTGTCTTGTCTATTCTTGCTGTTGATTATCCT 1155
Qy 1148 GTCGATAAGGTTGCTTGTTACGTATCTGATGATGGTGCTGCTATGCTTACTTTCGAAGCT 1207
Db 1156 GTCGATAAGGTTGCTTGTTACGTATCTGATGATGGTGCTGCTATGCTTACTTTCGAAGCT 1215
Qy 1208 CTTTCTGAGACCGCTGAATTCGCAAGGAAATGGGTTCCTTTCTGCAAGAAATATTGTATT 1267
Db 1216 CTTTCTGAGACCGCTGAATTCGCAAGGAAATGGGTTCCTTTCTGCAAGAAATATTGTATT 1275
Qy 1268 GAGCCTCGTGCTCCCGAATGGTATTTCTGCCATAAAATGGACTACTTGAAGAATAAAGTT 1327
Db 1276 GAGCCTCGTGCTCCCGAATGGTATTTCTGCCATAAAATGGACTACTTGAAGAATAAAGTT 1335
Qy 1328 CATCCCGCATTTGTTAGGGAGCGGCGAGCCATGAAGAGAGATTATGAAGAATTCAAAGTA 1387
Db 1336 CATCCCGCATTTGTTAGGGAGCGGCGAGCCATGAAGAGAGATTATGAAGAATTCAAAGTA 1395
Qy 1388 AAGATCAATGCTTTAGTAGCAACAGCACAGAAAGTGCCTGAGGATGGTTGGACTATGCAA 1447
Db 1396 AAGATCAATGCTTTAGTAGCAACAGCACAGAAAGTGCCTGAGGATGGTTGGACTATGCAA 1455
Qy 1448 GACGGTACACCTTGGCCCGGTAATAGTGTGCGAGATCATCCTGGCATGATTCAGGTCTTC 1507
Db 1456 GACGGTACACCTTGGCCCGGTAATAGTGTGCGAGATCATCCTGGCATGATTCAGGTCTTC 1515
Qy 1508 CTTGGAAGTGACGTTGTTCGTGATGTCGAAAACAACGAGTTGCCTCGATTAGTTTACGTT 1567
Db 1516 CTTGGAAGTGACGTTGTTCGTGATGTCGAAAACAACGAGTTGCCTCGATTAGTTTACGTT 1575
Qy 1568 TCTCGTGAGAAGACGCCCGGATTTGATCACCATAAGAAGGCTGGAGCTATGAATTCCCTG 1627
Db 1576 TCTCGTGAGAAGACGCCCGGATTTGATCACCATAAGAAGGCTGGAGCTATGAATTCCCTG 1635
Qy 1628 ATACGAGTCTCTGGGGTTCTATCAAATGCTCCTTACCTTCTGAATGTCGATTGTGATCAC 1687
Db 1636 ATACGAGTCTCTGGGGTTCTATCAAATGCTCCTTACCTTCTGAATGTCGATTGTGATCAC 1695
Qy 1688 TACATCAACAATAGCAAAGCTCTTAGAGAAGCAATGTGTTTCATGATGGATCCTCAGTCA 1747
Db 1696 TACATCAACAATAGCAAAGCTCTTAGAGAAGCAATGTGTTTCATGATGGATCCTCAGTCA 1755
Qy 1748 GGAAAGAAAATCTGTTATGTTCAGTTCCCTCAAAGGTTCGATGGGATTGATAGGCACGAT 1807
Db 1756 GGAAAGAAAATCTGTTATGTTCAGTTCCCTCAAAGGTTCGATGGGATTGATAGGCACGAT 1815
Qy 1808 CGATACTCAAATCGCAATGTTGTGTTCTTTGATATCAATATGAAAGGTTTGGATGGGCTA 1867
Db 1816 CGATACTCAAATCGCAATGTTGTGTTCTTTGATATCAATATGAAAGGTTTGGATGGGCTA 1875
Qy 1868 CAAGGGCCTATATACGTCGGTACAGGTTGTGTTTTCAGGAGGCAAGCGCTTTACGGATTT 1927
Db 1876 CAAGGGCCTATATACGTCGGTACAGGTTGTGTTTTCAGGAGGCAAGCGCTTTACGGATTT 1935
Qy 1928 GATGCACCGAAGAAGAAGAAGGGCCCACGTAAGACATGCAATTGCTGGCCAAAATGGTGT 1987
Db 1936 GATGCACCGAAGAAGAAGAAGGGCCCACGTAAGACATGCAATTGCTGGCCAAAATGGTGT 1995
Qy 1988 CTCCTATGTTTTGGTTCAAGAAAGAATCGTAAAGCAAAGACAGTGGCTGCGGATAAGAAG 2047
Db 1996 CTCCTATGTTTTGGTTCAAGAAAGAATCGTAAAGCAAAGACAGTGGCTGCGGATAAGAAG 2055
Qy 2048 AAGAAGAATAGGGAAGCGTCAAAGCAGATCCACGCATTAGAAAATATCGAAGAGGGCCGC 2107
Db 2056 AAGAAGAATAGGGAAGCGTCAAAGCAGATCCACGCATTAGAAAATATCGAAGAGGGCCGC 2115
Qy 2108 GGTCATAAAGTTCTTAACGTAGAACAGTCAACCGAGGCAATGCAAATGAAGTTGCAGAAG 2167
Db 2116 GTCACTAAAGGTTCTAACGTAGAACAGTCAACCGAGGCAATGCAAATGAAGTTGGAGAAG 2175
Qy 2168 AAATATGGGCAGTCTCCTGTATTTGTTGCATCTGCGCGTCTGGAGAATGGTGGGATGGCT 2227
Db 2176 AAATTTGGGCAGTCTCCTGTATTTGTTGCATCTGCGCGTATGGAGAATGGTGGGATGGCT 2235
Qy 2228 AGAAACGCAAGCCCGGCTTGTCTGCTTAAAGAAGCCATCCAAGTCATTAGTCGCGGATAT 2287
Db 2236 AGAAACGCAAGCCCGGCTTGTCTGCTTAAAGAAGCCATCCAAGTCATTAGTTGCGGATAT 2295
Qy 2288 GAAGATAAAACTGAATGGGGAAAAGAGATTGGGTGGATCTATGGTTCTGTTACCGAAGAT 2347
Db 2296 GAAGATAAAACTGAATGGGGAAAAGAGATTGGGTGGATCTATGGTTCTGTTACCGAAGAT 2355

```
Db       2296 GAAGATAAAACTGAATGGGGAAAAGAGATTGGGTGGATCTATGGTTCTGTTACCGAAGAT 2355

Qy       2348 ATTCTTACGGGTTCTAAGATGCATTCTCATGGTTGGAGACATGTTTATTGTACACCAAAG 2407
             |||||||||||||| |||||||||||||||||||||||||  |||||||||||||||||||
Db       2356 ATTCTTACGGGTTTTAAGATGCATTCTCATGGTTGGAGATCTGTTTATTGTACACCAAAG 2415

Qy       2408 TTAGCGGCTTTCAAAGGATCAGCTCCAATCAATCTTTCGGATCGTCTCCATCAAGTTCTT 2467
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2416 TTAGCGGCTTTCAAAGGATCAGCTCCAATCAATCTTTCGGATCGTCTCCATCAAGTTCTT 2475

Qy       2468 CGATGGGCGCTTGGGTCGGTTGAGATTTTCTTGAGTAGGCATTGTCCTATTTGGTATGGT 2527
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2476 CGATGGGCGCTTGGGTCGGTTGAGATTTTCTTGAGTAGGCATTGTCCTATTTGGTATGGT 2535

Qy       2528 TATGGAGGTGGGTTGAAATGGCTTGAGCGGTTGTCCTACATTAACTCTGTGGTTTACCCG 2587
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2536 TATGGAGGTGGGTTGAAATGGCTTGAGCGGTTGTCCTACATTAACTCTGTGGTTTACCCG 2595

Qy       2588 TGGACCTCTCTACCGCTCATCGTTTACTGTTCTCTCCCTGCCATCTGTCTTCTCACTGGA 2647
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2596 TGGACCTCTCTACCGCTCATCGTTTACTGTTCTCTCCCTGCCATCTGTCTTCTCACTGGA 2655

Qy       2648 AAATTCATCGTTCCCGAGATTAGCAACTATGCGAGTATCCTCTTCATGGCGCTCTTCTCG 2707
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2656 AAATTCATCGTTCCCGAGATTAGCAACTATGCGAGTATCCTCTTCATGGCGCTCTTCTCG 2715

Qy       2708 TCGATTGCAATAACGGGTATTCTCGAGATGCAATGGGGCAAAGTTGGGATCGATGATTGG 2767
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2716 TCGATTGCAATAACGGGTATTCTCGAGATGCAATGGGGCAAAGTTGGGATCGATGATTGG 2775

Qy       2768 TGGAGAAACGAACAGTTTTGGGTCATTGGAGGTGTTTCTGCGCATCTGTTTGCTCTCTTC 2827
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2776 TGGAGAAACGAACAGTTTTGGGTCATTGGAGGTGTTTCTGCGCATCTGTTTGCTCTCTTC 2835

Qy       2828 CAAGGTCTCCTCAAGGTTCTTGCTGGTGTCGACACTAACTTCACAGTCACATCAAAAGCA 2887
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2836 CAAGGTCTCCTCAAGGTTCTTGCTGGTGTCGACACTAACTTCACAGTCACATCAAAAGCA 2895

Qy       2888 GCTGATGATGGAGAGTTCTCTGACCTTTACCTCTTCAAATGGACTTCACTTCTCATCCCT 2947
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2896 GCTGATGATGGAGAGTTCTCTGACCTTTACCTCTTCAAATGGACTTCACTTCTCATCCCT 2955

Qy       2948 CCAATGACTCTACTCATCATAAACGTCATTGGAGTCATAGTCGGAGTCTCTGATGCCATC 3007
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2956 CCAATGACTCTACTCATCATAAACGTCATTGGAGTCATAGTCGGAGTCTCTGATGCCATC 3015

Qy       3008 AGCAATGGATACGACTCGTGGGGACCGCTTTTCGGAAGACTGTTCTTTGCACTTTGGGTC 3067
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3016 AGCAATGGATACGACTCGTGGGGACCGCTTTTCGGAAGACTGTTCTTTGCACTTTGGGTC 3075

Qy       3068 ATCATTCATCTTTACCCGTTCCTTAAAGGTTTGCTTGGGAAACAAGATAGAATGCCAACC 3127
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3076 ATCATTCATCTTTACCCGTTCCTTAAAGGTTTGCTTGGGAAACAAGATAGAATGCCAACC 3135

Qy       3128 ATTATTGTCGTCTGGTCCATCCTCCTGGCCTCGATTCTTACACTTCTTTGGGTCCGGGTT 3187
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3136 ATTATTGTCGTCTGGTCCATCCTCCTGGCCTCGATTCTTACACTTCTTTGGGTCCGGGTT 3195

Qy       3188 AATCCGTTTGTGGCGAAAGGCGGTCCTATTCTCGAGATCTGTGGTTTAGACTGCTTGTGA 3247
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3196 AATCCGTTTGTGGCGAAAGGCGGTCCTATTCTCGAGATCTGTGGTTTAGACTGCTTGTGA 3255


RESULT 2
US-09-938-842A-559
; Sequence 559, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
;  APPLICANT: Harper, Jeff
;  APPLICANT:  Kreps, Joel
;  APPLICANT:  Wang, Xun
;  APPLICANT:  Zhu, Tong
;  TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
;  TITLE OF INVENTION:  SAME, AND METHODS OF USE
;  FILE REFERENCE: SCRIP1300-3
;  CURRENT APPLICATION NUMBER: US/09/938,842A
;  CURRENT FILING DATE:  2001-08-24
;  PRIOR APPLICATION NUMBER: US 60/227,866
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: US 60/264,647
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/300,111
;  PRIOR FILING DATE: 2001-06-22
;  NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 559
;   LENGTH: 3255
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
US-09-938-842A-559

  Query Match             56.4%;  Score 1944;  DB 10;  Length 3255;
  Best Local Similarity   76.3%;  Pred. No. 0;
  Matches 2484;  Conservative    0;  Mismatches  725;  Indels   45;  Gaps    6;

Qy          8 CGGTTAATCGCCGGTTCTCACAACAGGAATGAGTTTGTCCTCATTAATGCCGATGAGAAT 67
              ||| | || || || |||||||||||| || || || || |||||||| |||||||||| |
Db         16 CGGCTCATTGCTGGCTCTCACAACAGAAACGAATTCGTTCTCATTAACGCCGATGAGAGT 75

Qy         68 GCCCGAATAAGATCAGTCCAAGAGCTGAGTGGACAGACATGTCAAATCTGCAGAGATGAG 127
              ||| ||||| ||||||| ||||| |||||||| || ||||||||||||||  |||||||
Db         76 GCCAGAATACGATCAGTACAAGAACTGAGTGGGCAAACATGTCAAATCTGTGGAGATGAA 135

Qy        128 ATCGAATTGACTGTTGATGGAGAACCGTTTGTGGCATGTAACGAATGTGCATTCCCTGTG 187
              |||||||| || |||    | || |  ||||| || || |||||||| |||||||| ||
Db        136 ATCGAATTAACGGTTAGCAGTGAGCTCTTTGTTGCTTGCAACGAATGCGCATTCCCGGTT 195

Qy        188 TGTAGACCTTGCTATGAGTACGAAAGACGAGAAGGCAATCAAGCTTGTCCACAGTGCAAA 247
              |||||||| |||||||||| ||| |  ||||||| |||||||||||||| |||||||||
Db        196 TGTAGACCATGCTATGAGTATGAACGTAGAGAAGGAAATCAAGCTTGTCCTCAGTGCAAA 255

Qy        248 ACCCGTTTCAAACGTCTTAAAGGAAGTCCAAGAGTTGA---AGGTGATGAAGAGGAAGAT 304
              || || | |||| |  |||||| |||||| | |||||    || |||||||||| |||||
Db        256 ACTCGATACAAAAGGATTAAAGGTAGTCCACGGGTTGATGGAGATGATGAAGAAGAAGAA 315

Qy        305 GACATTGATGATTTAGACAATGAGTTTGAGTATGGAAATAATGGGATTGGATTTGATCAG 364
              |||||||||||| | ||  |||||||||          || |||||| |    ||| ||
Db        316 GACATTGATGATCTTGAGTATGAGTTTG---------ATCATGGGATGGACCCTGAACAT 366

Qy        365 GTTTCTGAAGGTATGTCAATCTCTCGTCGCAACTCCGGTTTCCCACAATCTGATTTGGAT 424
              |  ||||||           || || |  ||| |||||      |    ||  ||||||
Db        367 GCCGCTGAAGCCGCACTCTCTTCACGCCTTAACACCGGT------CGTGGTGGATTGGAT 420

Qy        425 TCAGCTCCACCTGGCTCTCAGATTCCATTGCTGACTTACGGCGACGAGGACGTTGAGATT 484
              |||||||||||||||||||||||||  | |||||||  | || || || | ||| ||
Db        421 TCAGCTCCACCTGGCTCTCAGATTCCTCTTTTGACTTATTGTGATGAAGATGCTGATATG 480

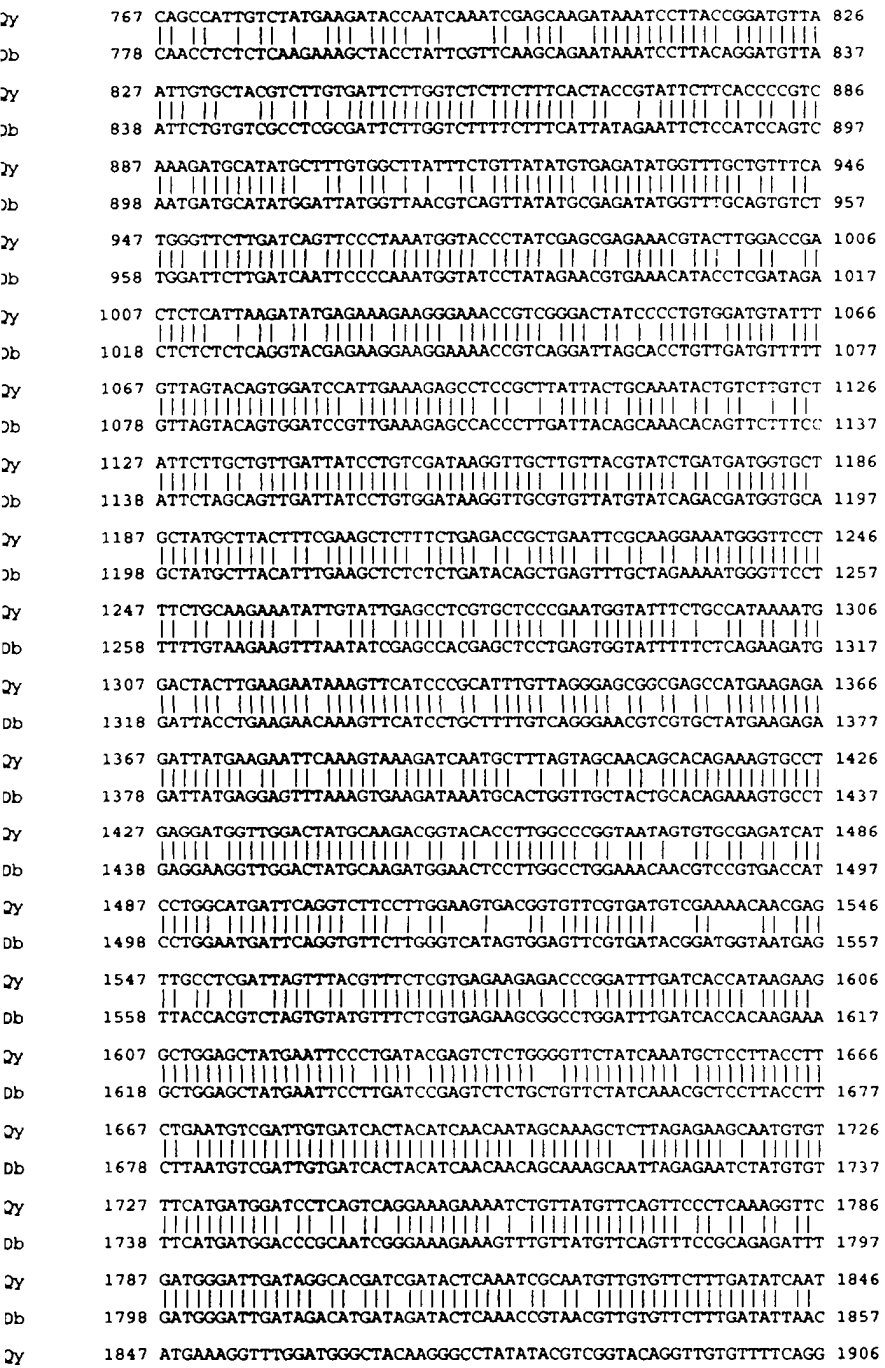
Qy        485 TCTTCTGATAGACATGCTCTTATTGTTCCTCCTTCACTTGGTGGTCATGGCAATAGAGTT 544
              | ||||||| | |||||||||| || |||||||||      ||  |||| ||| | ||
Db        481 TATTCTGATCGTCATGCTCTTATCGTGCCTCCTTCA---ACGGGATATGGGAATCGCGTC 537

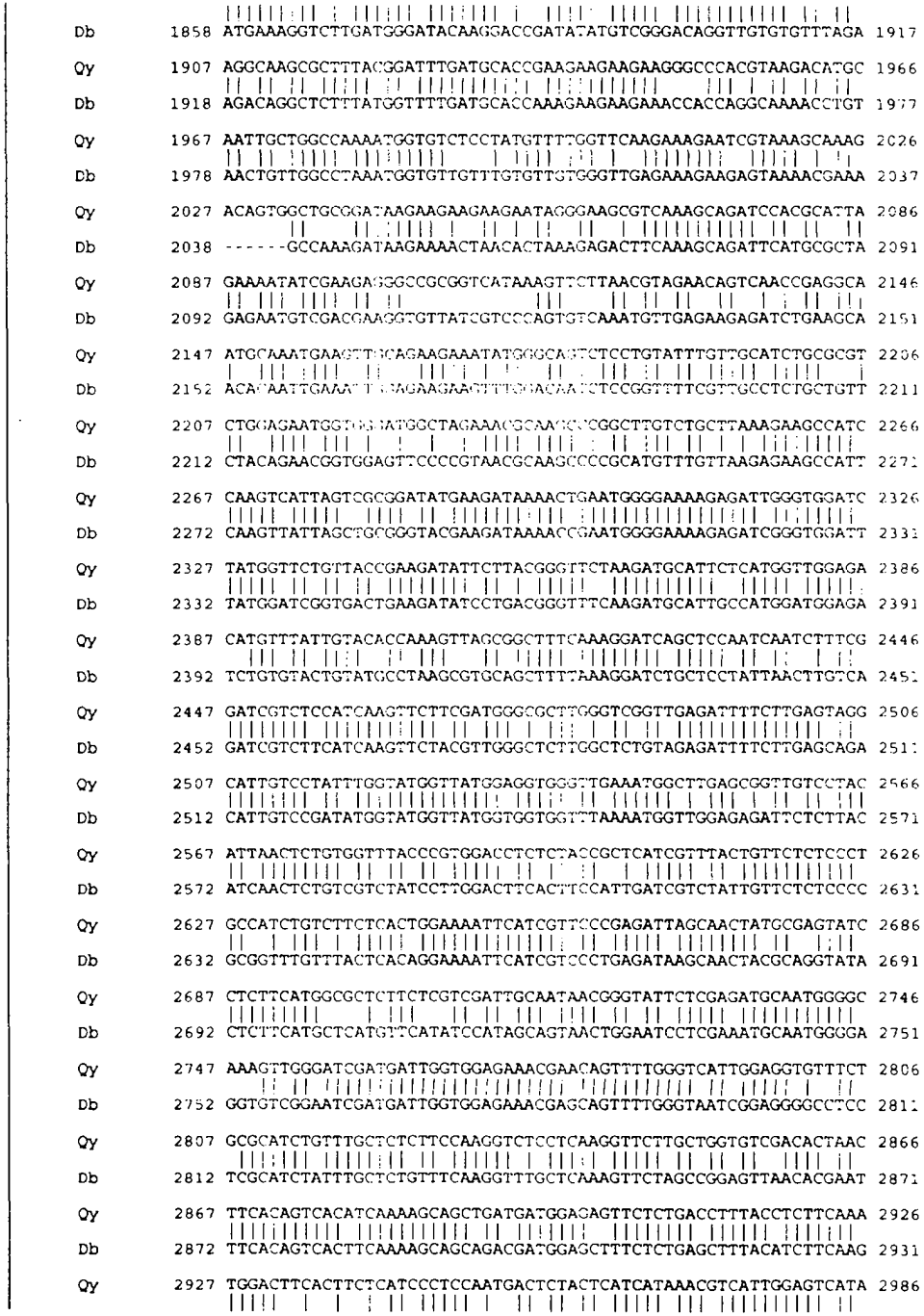
Qy        545 CATCCTGTTTCTCTTTCTGACCCGACCGTGGCTGCACATCGAAGGCTGATGGTACCTCAG 604
              ||||||   |  || | ||  |  | |   || ||||    ||    ||||| ||||||
Db        538 TATCCTGCACCGTTTACAGATTCTTCTGCACCTCCACAGGCGAGATCAATGGTTCCTCAG 597

Qy        605 AAAGATCTTGCGGTTTATGGTTATGGAAGTGTCGCTTGGAAAGATCGGATGGAGGAATGG 664
              |||||| ||||||  |||||||||||||||| |||||||| || || ||||| |  |||
Db        598 AAAGATATTGCGGAATATGGTTATGGAAGTGTTGCTTGGAAGGACCGTATGGAAGTTTGG 657

Qy        665 AAGAGAAAGCAGAATGAGAAACTTCAGGTTGTTAGGCATGAAGGAG-------------- 710
              ||||||  ||    || || ||||| ||  ||| ||||||||||
Db        658 AAGAGACGACAAGGCGAAAAGCTTCAAGTCATTAAGCATGAAGGAGGAAACAATGGTCGA 717

Qy        711 ----ATCCTGATTTTGAAGATGGTGATGATGCTGATTTTCCAATGATGGATGAGGGAAGG 766
                  ||||   || ||    |||||| |||  | || ||||||||||| |||||
Db        718 GGTTCCAATGATGACGACGAACTAGATGATCCTGACATGCCTATGATGGATGAAGGAAGA 777
```


Qy 767 CAGCCATTGTCTATGAAGATACCAATCAAATCGAGCAAGATAAATCCTTACCGGATGTTA 826
Db 778 CAACCTCTCTCAAGAAAGCTACCTATTCGTTCAAGCAGAATAAATCCTTACAGGATGTTA 837
Qy 827 ATTGTGCTACGTCTTGTGATTCTTGGTCTCTTCTTTCACTACCGTATTCTTCACCCCGTC 886
Db 838 ATTCTGTGTCGCCTCGCGATTCTTGGTCTTTTCTTTCATTATAGAATTCTCCATCCAGTC 897
Qy 887 AAAGATGCATATGCTTTGTGGCTTATTTCTGTTATATGTGAGATATGGTTTGCTGTTTCA 946
Db 898 AATGATGCATATGGATTATGGTTAACGTCAGTTATATGCGAGATATGGTTTGCAGTGTCT 957
Qy 947 TGGGTTCTTGATCAGTTCCCTAAATGGTACCCTATCGAGCGAGAAACGTACTTGGACCGA 1006
Db 958 TGGATTCTTGATCAATTCCCCAAATGGTATCCTATAGAACGTGAAACATACCTCGATAGA 1017
Qy 1007 CTCTCATTAAGATATGAGAAAGAAGGGAAACCGTCGGGACTATCCCCTGTGGATGTATTT 1066
Db 1018 CTCTCTCTCAGGTACGAGAAGGAAGGAAAACCGTCAGGATTAGCACCTGTTGATGTTTTT 1077
Qy 1067 GTTAGTACAGTGGATCCATTGAAAGAGCCTCCGCTTATTACTGCAAATACTGTCTTGTCT 1126
Db 1078 GTTAGTACAGTGGATCCGTTGAAAGAGCCACCCTTGATTACAGCAAACACAGTTCTTTCC 1137
Qy 1127 ATTCTTGCTGTTGATTATCCTGTCGATAAGGTTGCTTGTTACGTATCTGATGATGGTGCT 1186
Db 1138 ATTCTAGCAGTTGATTATCCTGTGGATAAGGTTGCGTGTTATGTATCAGACGATGGTGCA 1197
Qy 1187 GCTATGCTTACTTTCGAAGCTCTTTCTGAGACCGCTGAATTCGCAAGGAAATGGGTTCCT 1246
Db 1198 GCTATGCTTACATTTGAAGCTCTCTCTGATACAGCTGAGTTTGCTAGAAAATGGGTTCCT 1257
Qy 1247 TTCTGCAAGAAATATTGTATTGAGCCTCGTGCTCCCGAATGGTATTTCTGCCATAAAATG 1306
Db 1258 TTTTGTAAGAAGTTTAATATCGAGCCACGAGCTCCTGAGTGGTATTTTTCTCAGAAGATG 1317
Qy 1307 GACTACTTGAAGAATAAAGTTCATCCCGCATTTGTTAGGGAGCGGCGAGCCATGAAGAGA 1366
Db 1318 GATTACCTGAAGAACAAAGTTCATCCTGCTTTTGTCAGGGAACGTCGTGCTATGAAGAGA 1377
Qy 1367 GATTATGAAGAATTCAAAGTAAAGATCAATGCTTTAGTAGCAACAGCACAGAAAGTGCCT 1426
Db 1378 GATTATGAGGAGTTTAAAGTGAAGATAAATGCACTGGTTGCTACTGCACAGAAAGTGCCT 1437
Qy 1427 GAGGATGGTTGGACTATGCAAGACGGTACACCTTGGCCCGGTAATAGTGTGCGAGATCAT 1486
Db 1438 GAGGAAGGTTGGACTATGCAAGATGGAACTCCTTGGCCTGGAAACAACGTCCGTGACCAT 1497
Qy 1487 CCTGGCATGATTCAGGTCTTCCTTGGAAGTGACGGTGTTCGTGATGTCGAAAACAACGAG 1546
Db 1498 CCTGGAATGATTCAGGTGTTCTTGGGTCATAGTGGAGTTCGTGATACGGATGGTAATGAG 1557
Qy 1547 TTGCCTCGATTAGTTTACGTTTCTCGTGAGAAGAGACCCGGATTTGATCACCATAAGAAG 1606
Db 1558 TTACCACGTCTAGTGTATGTTTCTCGTGAGAAGCGGCCTGGATTTGATCACCACAAGAAA 1617
Qy 1607 GCTGGAGCTATGAATTCCCTGATACGAGTCTCTGGGGTTCTATCAAATGCTCCTTACCTT 1666
Db 1618 GCTGGAGCTATGAATTCCTTGATCCGAGTCTCTGCTGTTCTATCAAACGCTCCTTACCTT 1677
Qy 1667 CTGAATGTCGATTGTGATCACTACATCAACAATAGCAAAGCTCTTAGAGAAGCAATGTGT 1726
Db 1678 CTTAATGTCGATTGTGATCACTACATCAACAACAGCAAAGCAATTAGAGAATCTATGTGT 1737
Qy 1727 TTCATGATGGATCCTCAGTCAGGAAAGAAAATCTGTTATGTTCAGTTCCCTCAAAGGTTC 1786
Db 1738 TTCATGATGGACCCGCAATCGGGAAAGAAAGTTTGTTATGTTCAGTTTCCGCAGAGATTT 1797
Qy 1787 GATGGGATTGATAGGCACGATCGATACTCAAATCGCAATGTTGTGTTCTTTGATATCAAT 1846
Db 1798 GATGGGATTGATAGACATGATAGATACTCAAACCGTAACGTTGTGTTCTTTGATATTAAC 1857
Qy 1847 ATGAAAGGTTTGGATGGGCTACAAGGGCCTATATACGTCGGTACAGGTTGTGTTTTCAGG 1906


Db 1858 ATGAAAGGTCTTGATGGGATACAAGGACCGATATATGTCGGGACAGGTTGTGTGTTTAGA 1917
Qy 1907 AGGCAAGCGCTTTACGGATTTGATGCACCGAAGAAGAAGAAGGGCCCACGTAAGACATGC 1966
Db 1918 AGACAGGCTCTTTATGGTTTTGATGCACCAAAGAAGAAGAAACCACCAGGCAAAACCTGT 1977
Qy 1967 AATTGCTGGCCAAAATGGTGTCTCCTATGTTTTGGTTCAAGAAAGAATCGTAAAGCAAAG 2026
Db 1978 AACTGTTGGCCTAAATGGTGTTGTTTGTGTTGTGGGTTGAGAAAGAAGAGTAAAACGAAA 2037
Qy 2027 ACAGTGGCTGCGGATAAGAAGAAGAAGAATAGGGAAGCGTCAAAGCAGATCCACGCATTA 2086
Db 2038 ------GCCAAAGATAAGAAAACTAACACTAAAGAGACTTCAAAGCAGATTCATGCGCTA 2091
Qy 2087 GAAAATATCGAAGAGGGCCGCGGTCATAAAGTTCTTAACGTAGAACAGTCAACCGAGGCA 2146
Db 2092 GAGAATGTCGACGAAGGTGTTATCGTCCCAGTGTCAAATGTTGAGAAGAGATCTGAAGCA 2151
Qy 2147 ATGCAAATGAAGTTGCAGAAGAAATATGGGCAGTCTCCTGTATTTGTTGCATCTGCGCGT 2206
Db 2152 ACATAATTGAAATTGAGAAGAAGTTTGGACAATCTCCGGTTTTCGTTGCCTCTGCTGTT 2211
Qy 2207 CTGGAGAATGGTGGATGGCTAGAAACGCAAGCCCGGCTTGTCTGCTTAAAGAAGCCATC 2266
Db 2212 CTACAGAACGGTGGAGTTCCCCGTAACGCAAGCCCCGCATGTTTGTTAAGAGAAGCCATT 2271
Qy 2267 CAAGTCATTAGTCGCGGATATGAAGATAAAACTGAATGGGGAAAAGAGATTGGGTGGATC 2326
Db 2272 CAAGTTATTAGCTGCGGGTACGAAGATAAAACCGAATGGGGAAAAGAGATCGGGTGGATT 2331
Qy 2327 TATGGTTCTGTTACCGAAGATATTCTTACGGGTTCTAAGATGCATTCTCATGGTTGGAGA 2386
Db 2332 TATGGATCGGTGACTGAAGATATCCTGACGGGTTTCAAGATGCATTGCCATGGATGGAGA 2391
Qy 2387 CATGTTTATTGTACACCAAAGTTAGCGGCTTTCAAAGGATCAGCTCCAATCAATCTTTCG 2446
Db 2392 TCTGTGTACTGTATGCCTAAGCGTGCAGCTTTTAAAGGATCTGCTCCTATTAACTTGTCA 2451
Qy 2447 GATCGTCTCCATCAAGTTCTTCGATGGGCGCTTGGGTCGGTTGAGATTTTCTTGAGTAGG 2506
Db 2452 GATCGTCTTCATCAAGTTCTACGTTGGGCTCTTGGCTCTGTAGAGATTTTCTTGAGCAGA 2511
Qy 2507 CATTGTCCTATTTGGTATGGTTATGGAGGTGGGTTGAAATGGCTTGAGCGGTTGTCCTAC 2566
Db 2512 CATTGTCCGATATGGTATGGTTATGGTGGTGGTTTAAAATGGTTGGAGAGATTCTCTTAC 2571
Qy 2567 ATTAACTCTGTGGTTTACCCGTGGACCTCTCTACCGCTCATCGTTTACTGTTCTCTCCCT 2626
Db 2572 ATCAACTCTGTCGTCTATCCTTGGACTTCACTTCCATTGATCGTCTATTGTTCTCTCCCC 2631
Qy 2627 GCCATCTGTCTTCTCACTGGAAAATTCATCGTTCCCGAGATTAGCAACTATGCGAGTATC 2686
Db 2632 GCGGTTTGTTTACTCACAGGAAAATTCATCGTCCCTGAGATAAGCAACTACGCAGGTATA 2691
Qy 2687 CTCTTCATGGCGCTCTTCTCGTCGATTGCAATAACGGGTATTCTCGAGATGCAATGGGGC 2746
Db 2692 CTCTTCATGCTCATGTTCATATCCATAGCAGTAACTGGAATCCTCGAAATGCAATGGGGA 2751
Qy 2747 AAAGTTGGGATCGATGATTGGTGGAGAAACGAACAGTTTTGGGTCATTGGAGGTGTTTCT 2806
Db 2752 GGTGTCGGAATCGATGATTGGTGGAGAAACGAGCAGTTTTGGGTAATCGGAGGGGCCTCC 2811
Qy 2807 GCGCATCTGTTTGCTCTCTTCCAAGGTCTCCTCAAGGTTCTTGCTGGTGTCGACACTAAC 2866
Db 2812 TCGCATCTATTTGCTCTGTTTCAAGGTTTGCTCAAAGTTCTAGCCGGAGTTAACACGAAT 2871
Qy 2867 TTCACAGTCACATCAAAAGCAGCTGATGATGGAGAGTTCTCTGACCTTTACCTCTTCAAA 2926
Db 2872 TTCACAGTCACTTCAAAAGCAGCAGACGATGGAGCTTTCTCTGAGCTTTACATCTTCAAG 2931
Qy 2927 TGGACTTCACTTCTCATCCCTCCAATGACTCTACTCATCATAAACGTCATTGGAGTCATA 2986

```
Db       2932 TGGACAACTTTGTTGATTCCTCCGACAACACTTCTGATCATTAACATCATTGGAGTTATT 2991

Qy       2987 GTCGGAGTCTCTGATGCCATCAGCAATGGATACGACTCGTGGGGACCGCTTTTCGGAAGA 3046

Db       2992 GTCGGCGTTTCTGATGCCATTAGCAATGGCTATGACTCATGGGGACCTCTCTTTGGGAGA 3051

Qy       3047 CTGTTCTTTGCACTTTGGGTCATCATTCATCTTTACCCGTTCCTTAAAGGTTTGCTTGGG 3106

Db       3052 CTTTTCTTCGCTCTTTGGGTCATTGTTCATTTATACCCATTCCTCAAGGGAATGCTTGGG 3111

Qy       3107 AAACAAGATAGAATGCCAACCATTATTGTCGTCTGGTCCATCCTCCTGGCCTCGATTCTT 3166

Db       3112 AAGCAAGACAAAATGCCTACGATTATTGTGGTCTGGTCTATTCTTCTAGCTTCGATCTTG 3171

Qy       3167 ACACTTCTTTGGGTCCGGGTTAATCCGTTTGTGGCGAAAGGCGGTCCTATTCTCGAGATC 3226

Db       3172 ACACTCTTGTGGGTCAGAGTTAACCCGTTTGTGGCTAAAGGGGGACCAGTGTTGGAGATC 3231

Qy       3227 TGTGGTTTAGACTG 3240

Db       3232 TGTGGTCTGAATTG 3245


RESULT 3
US-10-229-193-7
; Sequence 7, Application US/10229193
; Publication No. US20030126643A1
; GENERAL INFORMATION:
;  APPLICANT: Arioli, Antonio Williamson, Richard E.
;  APPLICANT:  Betzner, Andreas S.
;  APPLICANT:  Peng, Liangcai
;  TITLE OF INVENTION: Manipulation of Cellulose and/or Beta-1.4-Glucan
;  FILE REFERENCE: 96-98A
;  CURRENT APPLICATION NUMBER: US/10/229,193
;  CURRENT FILING DATE:  2002-08-26
;  PRIOR APPLICATION NUMBER: 09/221,013
;  PRIOR FILING DATE: 1998-12-23
;  PRIOR APPLICATION NUMBER: PCT/AU97/00402
;  PRIOR FILING DATE: 1997-06-24
;  PRIOR APPLICATION NUMBER: AU PO0699
;  PRIOR FILING DATE: 1996-06-27
;  NUMBER OF SEQ ID NOS: 36
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
;  LENGTH: 3828
;  TYPE: DNA
;  ORGANISM: Arabidopsis thaliana
;  FEATURE:
;  NAME/KEY: CDS
;  LOCATION: (239)..(3490)
US 10-229-193-7

  Query Match           56.2%;  Score 1934.4;  DB 14;  Length 3828;
  Best Local Similarity  76.2%;  Pred. No. 0;
  Matches 2478;  Conservative    0;  Mismatches  731;  Indels   45;  Gaps    6;

Qy          8 CGGTTAATCGCCGGTTCTCACAACAGGAATGAGTTTGTCCTCATTAATGCCGATGAGAAT 67

Db        254 CGGCTCATTGCTGGCTCTCACAACAGAAACGAATTCGTTCTCATTAACGCCGATGAGAGT 313

Qy         68 GCCCGAATAAGATCAGTCCAAGAGCTGAGTGGACAGACATGTCAAATCTGCAGAGATGAG 127

Db        314 GCCAGAATACGATCAGTACAAGAACTGAGTGGGCAAACATGTCAAATCTGTGGAGATGAA 373

Qy        128 ATCGAATTGACTGTTGATGGAGAACCGTTTGTGGCATGTAACGAATGTGCATTCCCTGTG 187

Db        374 ATCGAATTAACGGTTAGCAGTGAGCTCTTTGTTGCTTGCAACGAATGCGCATTCCCGGTT 433

Qy        188 TGTAGACCTTGCTATGAGTACGAAAGACGAGAAGGCAATCAAGCTTGTCCACAGTGCAAA 247

Db        434 TGTAGACCATGCTATGAGTATGAACGTAGAGAAGGAAATCAAGCTTGTCCTCAGTGCAAA 493

Qy        248 ACCCGTTTCAAACGTCTTAAAGGAAGTCCAAGAGTTGA-- AGGTGATGAAGAGGAAGAT 304

Db        494 ACTCGATACAAAAGGATTAAAGGTAGTCCACGGGTTGATGGAGATGATGAAGAAGAAGAA 553

Qy        305 GACATTGATGATTTAGACAATGAGTTTGAGTATGGAAATAATGGGATTGGATTTGATCAG 364

Db        554 GACATTGATGATCTTGAGTATGAGTTTG---------ATCATGGGATGGACCCTGAACAT 604

Qy        365 GTTTCTGAAGGTATGTCAATCTCTCGTCGCAACTCCGGTTTCCCACAATCTGATTTGGAT 424

Db        605 GCCGCTGAAGCCGCACTCTCTTCACGCCTTAACACCGGT------CGTGGTGGATTGGAT 658

Qy        425 TCAGCTCCACCTGGCTCTCAGATTCCATTGCTGACTTACGGCGACGAGGACGTTGAGATT 484

Db        659 TCAGCTCCACCTGGCTCTCAGATTCCTCTTTTGACTTATTGTGATGAAGATGCTGATATG 718

Qy        485 TCTTCTGATAGACATGCTCTTATTGTTCCTCCTTCACTTGGTGGTCATGGCAATAGAGTT 544

Db        719 TATTCTGATCGTCATGCTCTTATCGTGCCTCCTTCA---ACGGGATATGGGAATCGCGTC 775

Qy        545 CATCCTGTTTCTCTTTCTGACCCGACCGTGGCTGCACATCGAAGGCTGATGGTACCTCAG 604

Db        776 TATCCTGCACCGTTTACAGATTCTTCTGCACCTCCACAGGCGAGATCAATGGTTCCTCAG 835

Qy        605 AAAGATCTTGCGGTTTATGGTTATGGAAGTGTCGCTTGGAAAGATCGGATGGAGGAATGG 664

Db        836 AAAGATATTGCGGAATATGGTTATGGAAGTGTTGCTTGGAAGGACCGTATGGAAGTTTGG 895

Qy        665 AAGAGAAAGCAGAATGAGAAACTTCAGGTTGTTAGGCATGAAGGAG-------------- 710

Db        896 AAGAGACGACAAGGCGAAAAGCTTCAAGTCATTAAGCATGAAGGAGGAAACAATGGTCGA 955

Qy        711 ----ATCCTGATTTTGAAGATGGTGATGATGCTGATTTTCCAATGATGGATGAGGGAAGG 766

Db        956 GGTTCCAATGATGACGACGAACTAGATGATCCTGACATGCCTATGATGGATGAAGGAAGA 1015

Qy        767 CAGCCATTGTCTATGAAGATACCAATCAAATCGAGCAAGATAAATCCTTACCGGATGTTA 826

Db       1016 CAACCTCTCTCAAGAAAGCTACCTATTCGTTCAAGCAGAATAAATCCTTACAGGATGTTA 1075

Qy        827 ATTGTGCTACGTCTTGTGATTCTTGGTCTCTTCTTTCACTACCGTATTCTTCACCCCGTC 886

Db       1076 ATTCTGTGTCGCCTCGCGATTCTTGGTCTTTTCTTTCATTATAGAATTCTCCATCCAGTC 1135

Qy        887 AAAGATGCATATGCTTTGTGGCTTATTTCTGTTATATGTGAGATATGGTTTGCTGTTTCA 946

Db       1136 AATGATGCATATGGATTATGGTTAACGTCAGTTATATGCGAAATATGGTTTGCAGTGTCT 1195

Qy        947 TGGGTTCTTGATCAGTTCCCTAAATGGTACCCTATCGAGCGAGAAACGTACTTGGACCGA 1006

Db       1196 TGGATTCTTGATCAATTCCCCAAATGGTATCCTATAGAACGTGAAACATACCTCGATAGA 1255

Qy       1007 CTCTCATTAAGATATGAGAAAGAAGGGAAACCGTCGGGACTATCCCCTGTGGATGTATTT 1066

Db       1256 CTCTCTCTCAGGTACGAGAAGGAAGGAAAACCGTCAGGATTAGCACCTGTTGATGTTTTT 1315

Qy       1067 GTTAGTACAGTGGATCCATTGAAAGAGCCTCCGCTTATTACTGCAAATACTGTCTTGTCT 1126

Db       1316 GTTAGTACAGTGGATCCGTTGAAAGAGCCCCCCTTGATTACAGCAAACACAGTTCTTTCC 1375

Qy       1127 ATTCTTGCTGTTGATTATCCTGTCGATAAGGTTGCTTGTTACGTATCTGATGATGGTGCT 1186

Db       1376 ATTCTAGCAGTTGATTATCCTGTGGATAAGGTTGCGTGTTATGTATCAAACAATGGTGCA 1435

Qy       1187 GCTATGCTTACTTTCGAAGCTCTTTCTGAGACCGCTGAATTCGCAAGGAAATGGGTTCCT 1246

Db       1436 GCTATGCTTACATTTGAAGCTCTCTCTGATACAGCTGATTTTGCTACAAAATGGGTTCCT 1495

Qy       1247 TTCTGCAAGAAATATTGTATTGAGCCTCGTGCTCCCGAATGGTATTTCTGCCATAAAATG 1306

Db       1496 TTTTGTAAGAAGTTTAATATCGAGCCACGAGCTCCTGAGTGGTATTTTTCTCAGAAGATG 1555

Qy       1307 GACTACTTGAAGAATAAAGTTCATCCCGCATTTGTTAGGGAGCGGCGAGCCATGAAGAGA 1366
```

```
Db       1556 GATTACCTGAAGAACAAAGTTCATCCTGCTTTTGTCAGGGAACGTCGTGCTATGAAGAGA 1615

Qy       1367 GATTATGAAGAATTCAAAGTAAAGATCAATGCTTTAGTAGCAACAGCACAGAAAGTGCCT 1426

Db       1616 GATTATGAAGAGTTTAAAGTGAAGATAAATGCACTGGTTGCTACTGCACAGAAAGTGCCT 1675

Qy       1427 GAGGATGGTTGGACTATGCAAGACGGTACACCTTGGCCCGGTAATAGTGTGCGAGATCAT 1486

Db       1676 GAGGAACGTTGGACTATGCAAGATGGAACTCCTTGGCCTGGAAACAACGTCCGTGACCAT 1735

Qy       1487 CCTGGCATGATTCAGGTCTTCCTTGGAAGTGACGGTGTTCGTGATGTCGAAAACAACGAG 1546

Db       1736 CCTGGAATGATTCAGGTGTTCTTGGGTCATAGTGGAGTTCGTGATACGGATGGTAATGAG 1795

Qy       1547 TTGCCTCGATTAGTTTACGTTTCTCGTGAGAAGAGACCCGGATTTGATCACCATAAGAAG 1606

Db       1796 TTACCACGTCTAGTGTATGTTTCTCGTGAGAAGCGGCCTGGATTTGATCACCACAAGAAA 1855

Qy       1607 GCTGGAGCTATGAATTCCCTGATACGAGTCTCTGGGGTTCTATCAAATGCTCCTTACCTT 1666

Db       1856 GCTGGAGCTATGAATTCCTTGATCCGAGTCTCTGCTGTTCTATCAAACGCTCCTTACCTT 1915

Qy       1667 CTGAATGTCGATTGTGATCACTACATCAACAATAGCAAAGCTCTTAGAGAAGCAATGTGT 1726

Db       1916 CTTAATGTCGATTGTGATCACTACATCAACAACAGCAAAGCAATTAGAGAATCTATGTGT 1975

Qy       1727 TTCATGATGGATCCTCAGTCAGGAAAGAAAATCTGTTATGTTCAGTTCCCTCAAAGGTTC 1786

Db       1976 TTCATGATGGACCCGCAATCGGGAAAGAAAGTTTGTTATGTTCAGTTTCCGCAGAGATTT 2035

Qy       1787 GATGGGATTGATAGGCACGATCGATACTCAAATCGCAATGTTGTGTTCTTTGATATCAAT 1846

b        2036 GATGGGATTGATAGACATGATAGATACTCAAACCGTAACGTTGTGTTCTTTGATATTAAC 2095

Qy       1847 ATGAAAGGTTTGGATGGGCTACAAGGGCCTATATACGTCGGTACAGGTTGTGTTTTCAGG 1906

Db       2096 ATGAAAGGTCTTGATGGGATACAAGGACCGATATATGTCGGGACAGGTTGTGTGTTTAGA 2155

Qy       1907 AGGCAAGCGCTTTACGGATTTGATGCACCGAAGAAGAAGAAGGGCCCACGTAAGACATGC 1966

Db       2156 AAACAGGCTCTTTATGGTTTTGATGCACCAAAGAAGAAGAAACCACCAGGCAAAACCTGT 2215

Qy       1967 AATTGCTGGCCAAAATGGTGTCTCCTATGTTTTGGTTCAAGAAAGAATCGTAAAGCAAAG 2026

Db       2216 AACTGTTGGCCTAAATGGTGTTGTTTGTGTTGTGGGTTGAGAAAGAAGAGTAAAACGAAA 2275

Qy       2027 ACAGTGGCTGCGGATAAGAAGAAGAAGAATAGGGAAGCGTCAAAGCAGATCCACGCATTA 2086

Db       2276 ------GCCACAGATAAGAAAACTAACACTAAAGAGACTTCAAAGCAGATTCATGCGCTA 2329

Qy       2087 GAAAATATCGAAGAGGGCCGCGGTCATAAAGTTCTTAACGTAGAACAGTCAACCGAGGCA 2146

Db       2330 GAGAATGTCGACGAAGGTGTTATCGTCCCAGTGTCAAATGTTGAGAAGAGATCTGAAGCA 2389

Qy       2147 ATGCAAATGAAGTTGCAGAAGAAATATGGGCAGTCTCCTGTATTTGTTGCATCTGCGCGT 2206

Db       2390 ACACAATTGAAATTGGAGAAGAAGTTTGGACAATCTCCGGTTTTCGTTGCCTCTGCTGTT 2449

Qy       2207 CTGGAGAATGGTGGGATGGCTAGAAACGCAAGCCCGGCTTGTCTGCTTAAAGAAGCCATC 2266

Db       2450 CTACAGAACGGTGGAGTTCCCCGTAACGCAAGCCCCGCATGTTTGTTAAGAGAAGCCATT 2509

Qy       2267 CAAGTCATTAGTCGCGGATATGAAGATAAAACTGAATGGGGAAAAGAGATTGGGTGGATC 2326

Db       2510 CAAGTTATTAGCTGCGGGTACCAAGATAAAACCGAATGGGGAAAAGAGATCGGGTGGATT 2569

Qy       2327 TATGGTTCTGTTACCGAAGATATTCTTACGGGTTCTAAGATGCATTCTCATGGTTGGAGA 2386

Db       2570 TATGGATCGGTGACTGAAGATATCCTGACGGGTTTCAAGATGCATTGCCATGGATGGAGA 2629

Qy       2387 CATGTTTATTGTACACCAAAGTTAGCGGCTTTCAAAGGATCAGCTCCAATCAATCTTTCG 2446

Db       2630 TCTGTGTACTGTATGCCTAAGCGTGCAGCTTTTAAAGGATCTGCTCCTATTAACTTGTCA 2689
```

```
Qy       2447 GATCGTCTCCATCAAGTTCTTCGATGGGCGCTTGGGTCGGTTGAGATTTTCTTGAGTAGG 2506

Db       2690 GATCGTCTTCATCAAGTTCTACGTTGGGCTCTTGGCTCTGTAGAGATTTTCTTGAGCAGA 2749

Qy       2507 CATTGTCCTATTTGGTATGGTTATGGAGGTGGGTTGAAATGGCTTGAGCGGTTGTCCTAC 2566

Db       2750 CATTGTCCGATATGGTATGGTTATGGTGGTGGTTTAAAATGGTTGGAGAGATTCTCTTAC 2809

Qy       2567 ATTAACTCTGTGGTTTACCCGTGGACCTCTCTACCGCTCATCGTTTACTGTTCTCTCCCT 2626

Db       2810 ATCAACTCTGTCGTCTATCCTTGGACTTCACTTCCATTGATCGTCTATTGTTCTCTCCCC 2869

Qy       2627 GCCATCTGTCTTCTCACTGGAAAATTCATCGTTCCCGAGATTAGCAACTATGCGAGTATC 2686

Db       2870 GCGGTTTGTTTACTCACAGGAAAATTCATCGTCCCTGAGATAAGCAACTACGCAGGTATA 2929

Qy       2687 CTCTTCATGGCGCTTTTCTCGTCGATTGCAATAACGGGTATTCTCGAGATGCAATGGGGC 2746

Db       2930 CTCTTCATGCTCATGTTCATATCCATAGCAGTAACTGGAATCCTCGAAATGCAATGGGGA 2989

Qy       2747 AAAGTTGGGATCGATGATTGGTGGAGAAACGAACAGTTTTGGGTCATTGGAGGTGTTTCT 2806

Db       2990 GGTGTCGGAATCGATGATTGGTGGAGAAACGAGCAGTTTTGGGTAATCGGAGGGCCTCC 3049

Qy       2807 GCGCATCTGTTTGCTCTCTTCCAAGGTCTCCTCAAGGTTCTTGCTGGTGTCGACACTAAC 2866

Db       3050 TCGCATCTATTTGCTCTGTTTCAAGGTTTGCTCAAAGTTCTAGCCGGAGTTAACACGAAT 3109

Qy       2867 TTCACAGTCACATCAAAAGCAGCTGATGATGGAGAGTTCTCTGACCTTTACCTCTTCAAA 2926

Db       3110 TTCACAGTCACTTCAAAAGCAGCAGACGATGGAGCTTTCTCTGAGCTTTACATCTTCAAG 3169

Qy       2927 TGGACTTCACTTCTCATCCCTCCAATGACTCTACTCATCATAAACGTCATTGGAGTCATA 2986

Db       3170 TGGACAACTTTGTTGATTCCTCCGACAACACTTCTGATCATTAACATCATTGGAGTTATT 3229

Qy       2987 GTCGGAGTCTCTGATGCCATCAGCAATGGATACGACTCGTGGGGACCGCTTTTCGGAAGA 3046

Db       3230 GTCGGCGTTTCTGATGCCATTAGCAATGGCTATGACTCATGGGGACCTCTCTTTGGGAGA 3289

Qy       3047 CTGTTCTTTGCACTTTGGGTCATCATTCATCTTTACCCGTTCCTTAAAGGTTTGCTTGGG 3106

Db       3290 CTTTTCTTCGCTCTTTGGGTCATTGTTCATTTATACCCATTCCTCAAGGGAATGCTTGGG 3349

Qy       3107 AAACAAGATAGAATGCCAACCATTATTGTCGTCTGGTCCATCCTCCTGGCCTCGATTCTT 3166

Db       3350 AAGCAAGACAAAATGCCTACGATTATTGTGGTCTGGTCTATTCTTCTAGCTTCGATCTTG 3409

Qy       3167 ACACTTCTTTGGGTCCGGGTTAATCCGTTTGTGGCGAAAGGCGGTCCTATTCTCGAGATC 3226

Db       3410 ACACTCTTGTGGGTCAGAATTAACCCGTTTGTGGCTAAAGGGGGACCAGTGTTGGAGATC 3469

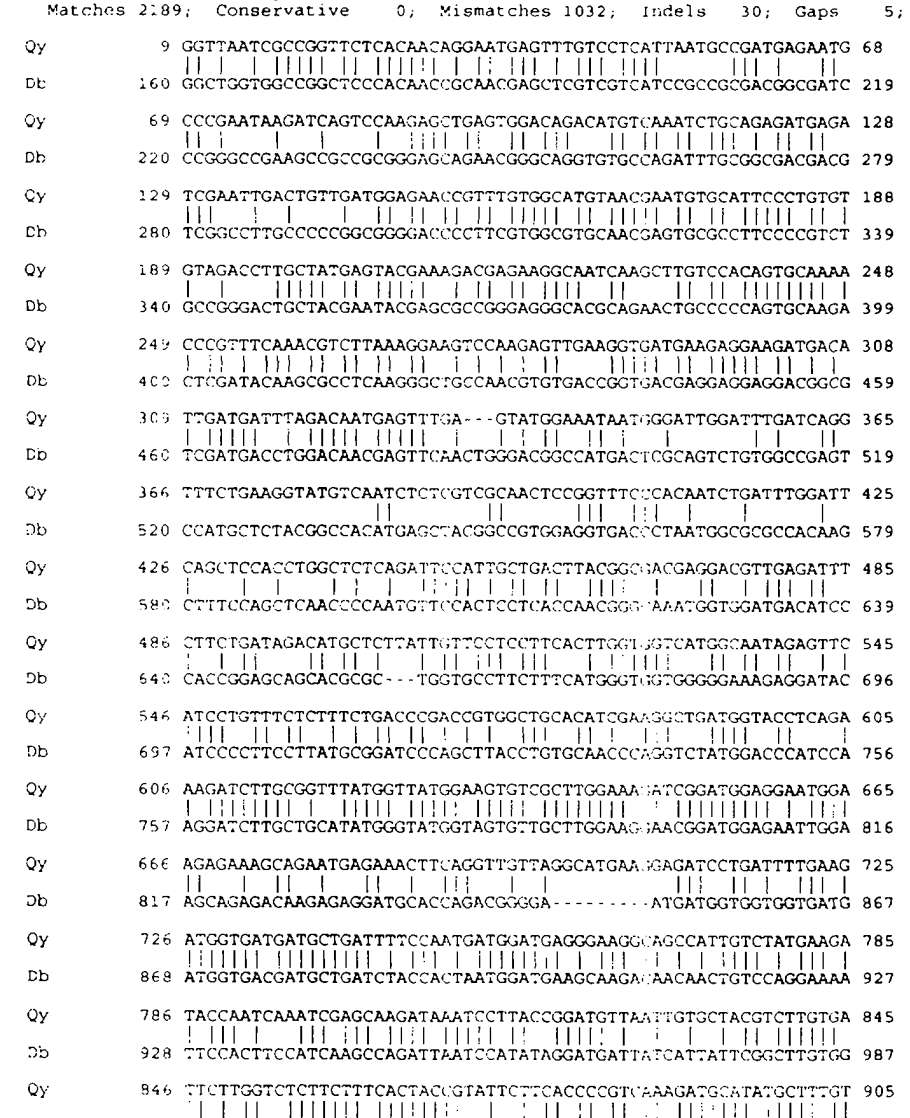
Qy       3227 TGTGGTTTAGACTG 3240

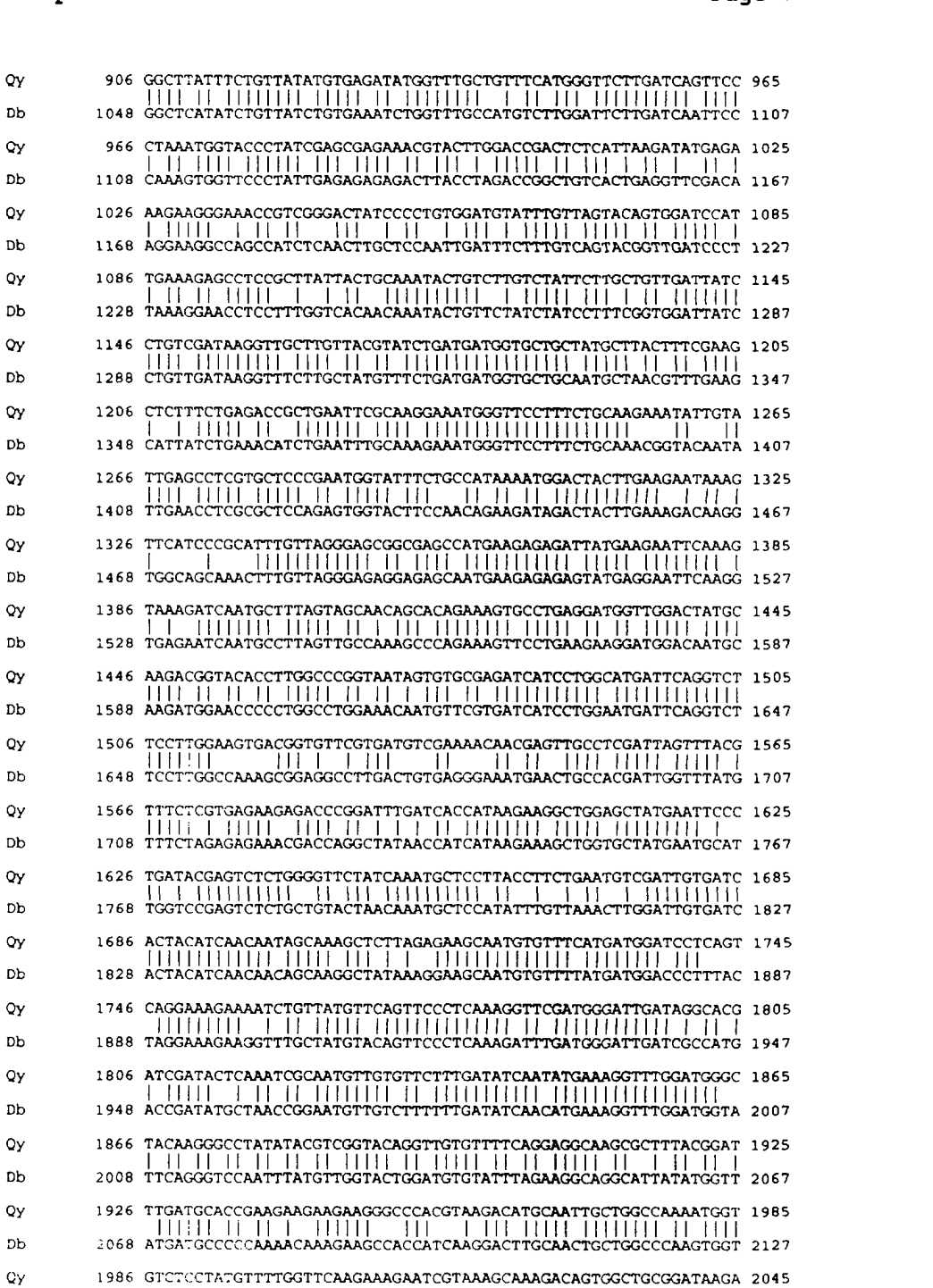
Db       3470 TGTGGTCTGAATTG 3483
```

```
RESULT 4
US-09-900-237-9
; Sequence 9, Application US/09900237
; Patent No. US20020120124A1
; GENERAL INFORMATION:
;  APPLICANT: Allen, Stephen
;  TITLE OF INVENTION: Plant Cellulose Synthases
;  FILE REFERENCE: BB1170 US CIP
;  CURRENT APPLICATION NUMBER: US/09/900,237
;  CURRENT FILING DATE:  2001-07-06
;  PRIOR APPLICATION NUMBER: 60/092,844
;  PRIOR FILING DATE: 1998-07-14
;  PRIOR APPLICATION NUMBER: PCT/US99/15871
;  PRIOR FILING DATE: 1999-07-13
;  PRIOR APPLICATION NUMBER: 09/720383
```

; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 9
; LENGTH: 3936
; TYPE: DNA
; ORGANISM: Zea mays
US-09-900-237-9

Query Match 43.3%; Score 1489.8; DB 10; Length 3936;
Best Local Similarity 67.3%; Pred. No. 0;
Matches 2189; Conservative 0; Mismatches 1032; Indels 30; Gaps 5;





```
b    2128 GCTTTTGCTGTTGCTGCTTTGGCAATAGGAAGCAAAAGAAGACTACCAAACCCAAAACAG 2187
y    2046 AGAAGAAGAA----------TAGGGAAGCGTCAAAGCAGATCCACGCATTAGAAAATA 2093
b    2188 AGAAGAAAAAGTTATTATTTTTCAAGAAAGAAGAGAACCAATCCCCTGCATATGCTCTTG 2247
y    2094 TCGAAGAGGGCCGCGGTCATAAAGTTCTTAACGTAGAACAGTCAACCGAGGCAATGCAAA 2153
b    2248 GTGAAATTGACGAAGCTGCTCCAGGAGCTGAGAATGAAAAGGCCGGTATTGTAAATCAAC 2307
y    2154 TGAAGTTGCAGAAGAAATATGGGCAGTCTCCTGTATTTGTTGCATCTGCGCGTCTGGAGA 2213
b    2308 AAAAATTAGAAAAGAAATTTGGCCAATCTTCTGTTTTTGTTACATCCACACTTCTCGAGA 2367
y    2214 ATGGTGGGATGGCTAGAAACGCAAGCCCGGCTTGTCTGCTTAAAGAAGCCATCCAAGTCA 2273
b    2368 ATGGTGGAACCTTGAAGAGTGCAAGTCCTGCTTCTCTTTTGAAAGAAGCTATACATGTCA 2427
y    2274 TTAGTCGCGGATATGAAGATAAAACTGAATGGGGAAAAGAGATTGGGTGGATCTATGGTT 2333
b    2428 TTAGTTGTGGTTATGAAGACAAGACAGACTGGGGAAAAGAGATTGGCTGGATCTATGGAT 2487
y    2334 CTGTTACCGAAGATATTCTTACGGGTTCTAAGATGCATTCTCATGGTTGGAGACATGTTT 2393
b    2488 CAGTTACAGAAGATATTCTAACTGGTTTCAAGATGCATTGTCATGGTTGGCGGTCAATTT 2547
y    2394 ATTGTACACCAAAGTTAGCGGCTTTCAAAGGATCAGCTCCAATCAATCTTTCGGATCGTC 2453
b    2548 ACTGCATACCTAAACGGGTTGCATTCAAAGGTTCTGCACCTCTGAATCTTTCAGATCGTC 2607
y    2454 TCCATCAAGTTCTTCGATGGGCGCTTGGGTCGGTTGAGATTTTCTTGAGTAGGCATTGTC 2513
b    2608 TTCACCAGGTGCTTCGGTGGGCTCTTGGGTCTATTGAGATCTTCTTCAGCAATCATTGCC 2667
y    2514 CTATTTGGTATGGTTATGGAGGTGGGTTGAAATGGCTTGAGCGGTTGTCCTACATTAACT 2573
b    2668 CTCTTTGGTATGGGTATGGTGGCGGTCTGAAATTTTTGGAAAGATTTTCCTACATCAACT 2727
y    2574 CTGTGGTTTACCCGTGGACCTCTCTACCGCTCATCGTTTACTGTTCTCTCCCTGCCATCT 2633
b    2728 CCATCGTGTATCCTTGGACATCTATTCCCCTCTTGGCTTACTGTACATTGCCTGCCATCT 2787
y    2634 GTCTTCTCACTGGAAAATTCATCGTTCCCGAGATTAGCAACTATGCGAGTATCCTCTTCA 2693
b    2788 GTTTATTGACAGGGAAATTTATCACTCCAGAGCTGAATAATGTTGCCAGCCTGTGGTTCA 2847
y    2694 TGGCGCTCTTCTCGTCGATTGCAATAACGGGTATTCTCGAGATGCAATGGGGCAAAGTTG 2753
b    2848 TGTCACTTTTTATCTGCATTTTTGCTACGAGCATCCTAGAAATGAGATGGAGTGGTGTTG 2907
y    2754 GGATCGATGATTGGTGGAGAAACGAACAGTTTTGGGTCATTGGAGGTGTTTCTGCGCATC 2813
b    2908 GAATTGATGACTGGTGGAGGAATGAGCAGTTCTGGGTCATTGGAGGTGTGTCCTCACACC 2967
y    2814 TGTTTGCTCTCTTCCAAGGTCTCCTCAAGGTTCTTGCTGGTGTCGACACTAACTTCACAG 2873
b    2968 TCTTTGCTGTGTTCCAGGGACTTCTCAAGGTCATAGCTGGTGTTGATACAAGCTTCACCG 3027
y    2874 TCACATCAAAAGCAGCTGATGATGGAGAGTTCTCTGACCTTTACCTCTTCAAATGGACTT 2933
b    3028 TGACATCAAAGGGTGGAGATGATGAGGAGTTCTCAGAGCTATATACATTCAAATGGACTA 3087
y    2934 CACTTCTCATCCCTCCAATGACTCTACTCATCATAAACGTCATTGGAGTCATAGTCGGAG 2993
b    3088 CCTTATTGATACCTCCTACCACCTTGCTTCTATTGAACTTCATTGGTGTGGTCGCTGGCG 3147
y    2994 TCTCTGATGCCATCAGCAATGGATACGACTCGTGGGGACCGCTTTTCGGAAGACTGTTCT 3053
b    3148 TTTCAAATGCGATCAATAACGGATATGAGTCATGGGGCCCCCTCTTTGGGAAGCTATTCT 3207
y    3054 TTGCACTTTGGGTCATCATTCATCTTTACCCGTTCCTTAAAGGTTTGCTTGGGAAACAAG 3113
```

```
Db   3208 TTTCATTTTGGGTCATTGTCCATCTTTATCCTTTTCTCAAAGGTTTGGTTGGAAGGCAAA 3267

Qy   3114 ATAGAATGCCAACCATTATTGTCGTCTGGTCCATCCTCCTGGCCTCGATTCTTACACTTC 3173

Db   3268 ACAGGACACCAACCATTGTCATCGTTTGGTCCATTCTGCTGGCTTCAATCTTCTCGCTCC 3327

Qy   3174 TTTGGGTCCGGGTTAATCCGTTTGTGGCGA---AAGGCGGTCCTATTCTCGAGATCTGTG 3230

Db   3328 TTTGGGTTCGGATTGATCCTTTCCTTGCGAAGGATGATGGTCCGCTTCTTGAGGAGTGTG 3387

Qy   3231 GTTTAGACTGC 3241

Db   3388 GTTTGGATTGC 3398
```

RESULT 5
US-10-209-059-49
; Sequence 49, Application US/10209059
; Publication No. US20030163838A1
; GENERAL INFORMATION:
; APPLICANT: Dhugga, Kanwarpal S.
; APPLICANT: Wang, Haiyin
; TITLE OF INVENTION: Maize Cellulose Synthases and Uses
; TITLE OF INVENTION: Thereof
; FILE REFERENCE: 0864R2
; CURRENT APPLICATION NUMBER: US/10/209,059
; CURRENT FILING DATE: 2002-07-31
; PRIOR APPLICATION NUMBER: 60/096,822
; PRIOR FILING DATE: 1998-08-17
; PRIOR APPLICATION NUMBER: 09/371,383
; PRIOR FILING DATE: 1999-08-06
; PRIOR APPLICATION NUMBER: 09/550,483
; PRIOR FILING DATE: 2000-04-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 49
; LENGTH: 3969
; TYPE: DNA
; ORGANISM: Zea mays
US-10-209-059-49

Query Match 43.3%; Score 1489.8; DB 12; Length 3969;
Best Local Similarity 67.3%; Pred. No. 0;
Matches 2189; Conservative 0; Mismatches 1032; Indels 30; Gaps 5;

```
Qy      9 GGTTAATCGCCGGTTCTCACAACAGGAATGAGTTTGTCCTCATTAATGCCGATGAGAATG 68

Db    160 GGCTGGTGGCCGGCTCCCACAACCGCAACGAGCTCGTCGTCATCCGCCGCGACGGCGATC 219

Qy     69 CCCGAATAAGATCAGTCCAAGAGCTGAGTGGACAGACATGTCAAATCTGCAGAGATGAGA 128

Db    220 CCGGGCCGAAGCCGCCGCGGGAGCAGAACGGGCAGGTGTGCCAGATTTGCGGCGACGACG 279

Qy    129 TCGAATTGACTGTTGATGGAGAACCGTTTGTGGCATGTAACGAATGTGCATTCCCTGTGT 188

Db    280 TCGGCCTTGCCCCCGGCGGGGACCCCTTCGTGGCGTGCAACGAGTGCGCCTTCCCCGTCT 339

Qy    189 GTAGACCTTGCTATGAGTACGAAAGACGAGAAGGCAATCAAGCTTGTCCACAGTGCAAAA 248

Db    340 GCCGGGACTGCTACGAATACGAGCGCCGGGAGGGCACGCAGAACTGCCCCCAGTGCAAGA 399

Qy    249 CCCGTTTCAAACGTCTTAAAGGAAGTCCAAGAGTTGAAGGTGATGAAGAGGAAGATGACA 308

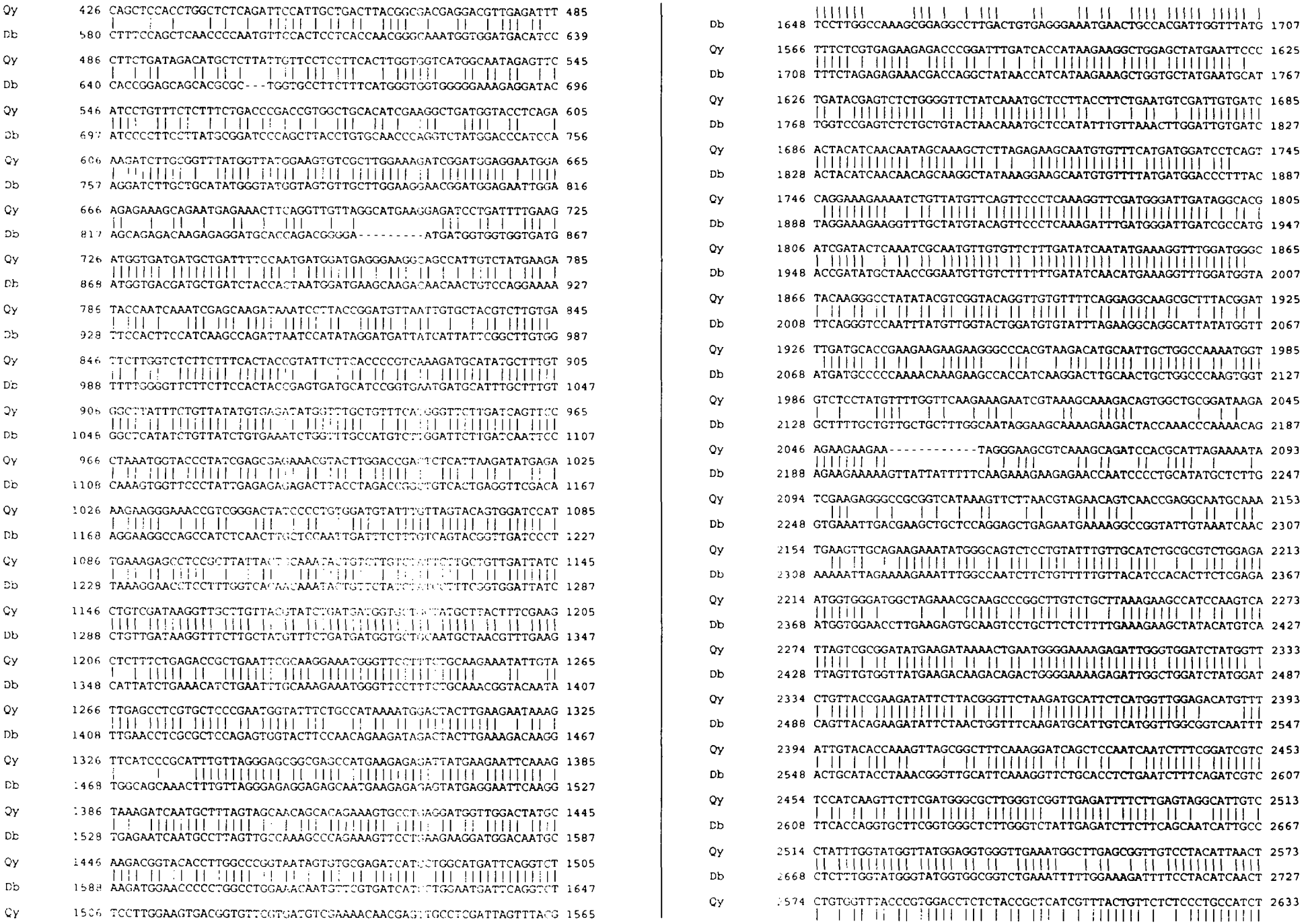
Db    400 CTCGATACAAGCGCCTCAAGGGCTGCCAACGTGTGACCGGTGACGAGGAGGAGGACGGCG 459

Qy    309 TTGATGATTTAGACAATGAGTTTGA---GTATGGAAATAATGGGATTGGATTTGATCAGG 365

Db    460 TCGATGACCTGGACAACGAGTTCAACTGGGACGGCCATGACTCGCAGTCTGTGGCCGAGT 519

Qy    366 TTTCTGAAGGTATGTCAATCTCTCGTCGCAACTCCGGTTTCCCACAATCTGATTTGGATT 425

Db    520 CCATGCTCTACGGCCACATGAGCTACGGCCGTGGAGGTGACCCTAATGGCGCGCCACAAG 579
```



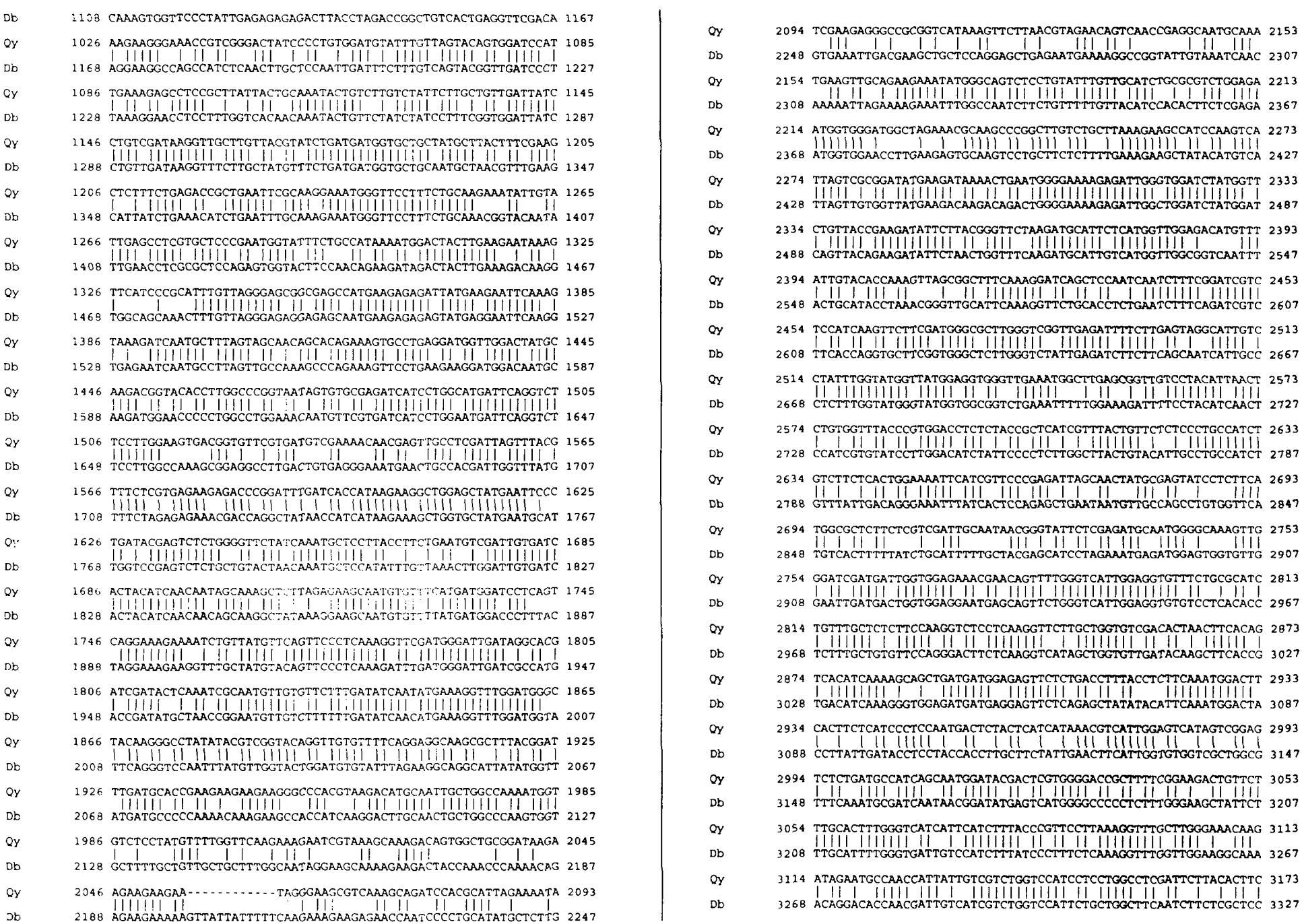
```
Db       2728 CCATCGTGTATCCTTGGACATCTATTCCCCTCTTGGCTTACTGTACATTGCCTGCCATCT 2787
Qy       2634 GTCTTCTCACTGGAAAATTCATCGTTCCCGAGATTAGCAACTATGCGAGTATCCTCTTCA 2693
Db       2788 GTTTATTGACAGGGAAATTTATCACTCCAGAGCTGAATAATGTTGCCAGCCTGTGGTTCA 2847
Qy       2694 TGGCGCTCTTCTCGTCGATTGCAATAACGGGTATTCTCGAGATGCAATGGGGCAAAGTTG 2753
Db       2848 TGTCACTTTTTATCTGCATTTTTGCTACGAGCATCCTAGAAATGAGATGGAGTGGTGTTG 2907
Qy       2754 GGATCGATGATTGGTGGAGAAACGAACAGTTTTGGGTCATTGGAGGTGTTTCTGCGCATC 2813
Db       2908 GAATTGATGACTGGTGGAGGAATGAGCAGTTCTGGGTCATTGGAGGTGTGTCCTCACACC 2967
Qy       2814 TGTTTGCTCTCTTCCAAGGTCTCCTCAAGGTTCTTGCTGGTGTCGACACTAACTTCACAG 2873
Db       2968 TCTTTGCTGTGTTCCAGGGACTTCTCAAGGTCATAGCTGGTGTTGATACAAGCTTCACCG 3027
Qy       2874 TCACATCAAAAGCAGCTGATGATGGAGAGTTCTCTGACCTTTACCTCTTCAAATGGACTT 2933
Db       3028 TGACATCAAAGGGTGGAGATGATGAGGAGTTCTCAGAGCTATATACATTCAAATGGACTA 3087
Qy       2934 CACTTCTCATCCCTCCAATGACTCTACTCATCATAAACGTCATTGGAGTCATAGTCGGAG 2993
Db       3088 CCTTATTGATACCTCCTACCACCTTGCTTCTATTGAACTTCATTGGTGTGGTCGCTGGCG 3147
Qy       2994 TCTCTGATGCCATCAGCAATGGATACGACTCGTGGGGACCGCTTTTCGGAAGACTGTTCT 3053
Db       3148 TTTCAAATGCGATCAATAACGGATATGAGTCATGGGGCCCCCTCTTTGGGAAGCTATTCT 3207
Qy       3054 TTGCACTTTGGGTCATCATTCATCTTTACCCGTTCCTTAAAGGTTTGCTTGGGAAACAAG 3113
Db       3208 TTGCATTTTGGGTGATTGTCCATCTTTATCCCTTTCTCAAAGGTTTGGTTGGAAGGCAAA 3267
Qy       3114 ATAGAATGCCAACCATTATTGTCGTCTGGTCCATCCTCCTGGCCTCGATTCTTACACTTC 3173
Db       3268 ACAGGACACCAACGATTGTCATCGTCTGGTCCATTCTGCTGGCTTCAATCTTCTCGCTCC 3327
Qy       3174 TTTGGGTCCGGGTTAATCCGTTTGTGGCGA---AAGGCGGTCCTATTCTCGAGATCTGTG 3230
Db       3328 TTTGGGTTCGGATTGATCCTTTCCTTGCGAAGGATGATGGTCCGCTTCTTGAGGAGTGTG 3387
Qy       3231 GTTTAGACTGC 3241
Db       3388 GTTTGGATTGC 3398
```

ESULT 6
JS-10-160-719-17
Sequence 17, Application US/10160719
Publication No. US20030167528A1
GENERAL INFORMATION:
APPLICANT: Kanwarpal S. Dhugga
APPLICANT: Timothy G. Helentjaris
APPLICANT: Benjamin A. Bowen
APPLICANT: Xun Wang
TITLE OF INVENTION: Maize Cellulose Synthases and Uses
TITLE OF INVENTION: Thereof
FILE REFERENCE: 0864
CURRENT APPLICATION NUMBER: US/10/160,719
CURRENT FILING DATE: 2002-06-03
PRIOR APPLICATION NUMBER: US/09/371,383A
PRIOR FILING DATE: 1999-08-06
PRIOR APPLICATION NUMBER: US 60/096,822
PRIOR FILING DATE: 1998-08-17
NUMBER OF SEQ ID NOS: 60
SOFTWARE: FastSEQ for Windows Version 3.0
SEQ ID NO 17
LENGTH: 3969
TYPE: DNA
ORGANISM: Zea mays
FEATURE:
NAME/KEY: CDS

; LOCATION: (144)...(3401)
US-10-160-719-17

Query Match 43.3%; Score 1489.8; DB 12; Length 3969;
Best Local Similarity 67.3%; Pred. No. 0;
Matches 2189; Conservative 0; Mismatches 1032; Indels 30; Gaps 5;

```
Qy          9 GGTTAATCGCCGGTTCTCACAACAGGAATGAGTTTGTCCTCATTAATGCCGATGAGAATG 68
Db        160 GGCTGGTGGCCGGCTCCCACAACCGCAACGAGCTCGTCGTCATCCGCCGCGACGGCGATC 219
Qy         69 CCCGAATAAGATCAGTCCAAGAGCTGAGTGGACAGACATGTCAAATCTGCAGAGATGAGA 128
Db        220 CCGGGCCGAAGCCGCCGCGGGAGCAGAACGGGCAGGTGTGCCAGATTTGCGGCGACGACG 279
Qy        129 TCGAATTGACTGTTGATGGAGAACCGTTTGTGGCATGTAACGAATGTGCATTCCCTGTGT 188
Db        280 TCGGCCTTGCCCCCGGCGGGGACCCCTTCGTGGCGTGCAACGAGTGCGCCTTCCCCGTCT 339
Qy        189 GTAGACCTTGCTATGAGTACGAAAGACGAGAAGGCAATCAAGCTTGTCCACAGTGCAAAA 248
Db        340 GCCGGGACTGCTA GAATACGAGCGCCGGGAGGGCACGCAGAACTGCCCCCAGTGCAAGA 399
Qy        249 CCCGTTTCAAACGTCTTAAAGGAAGTCCAAGAGTTGAAGGTGATGAAGAGGAAGATGACA 308
Db        400 CTCGATACAAGCGCCTCAAGGGCTGCCAACGTGTGACCGGTGACGAGGAGGAGGACGGCG 459
Qy        309 TTGATGATTTAGACAATGAGTTTGA---GTATGGAAATAATGGGATTGGATTTGATCAGG 365
Db        460 TCGATGACCTGGACAACGAGTTCAACTGGGACGGCCATGACTCGCAGTCTGTGGCCGAGT 519
Qy        366 TTTCTGAAGGTATGTCAATCTCTCGTCGCAACTCCGGTTTCCCACAATCTGATTTGGATT 425
Db        520 CCATGCTCTACGGCCACATGAGCTACGGCCGTGGAGGTGACCCTAATGGCGCGCCACAAG 579
Qy        426 CAGCTCCACCTGGCTCTCAGATTCCATTGCTGACTTACGGCGACGAGGACGTTGAGATTT 485
Db        580 CTTTCCAGCTCAACCCCAATGTTCCACTCCTCACCAACGGGCAAATGGTGGATGACATCC 639
Qy        486 CTTCTGATAGACATGCTCTTATTGTTCCTCCTTCACTTGGTGGTCATGGCAATAGAGTTC 545
Db        640 CACCGGAGCAGCACGCGC---TGGTGCCTTCTTTCATGGGTGGTGGGGGAAAGAGGATAC 696
Qy        546 ATCCTGTTTCTCTTTCTGACCCGACCGTGGCTGCACATCGAAGGCTGATGGTACCTCAGA 605
Db        697 ATCCCCTTCCTTATGCGGATCCCAGCTTACCTGTGCAACCCAGGTCTATGGACCCATCCA 756
Qy        606 AAGATCTTGCGGTTTATGGTTATGGAAGTGTCGCTTGGAAAGATCGGATGGAGGAATGGA 665
Db        757 AGGATCTTGCTGCATATGGGTATGGTAGTGTTGCTTGGAAGGAACGGATGGAGAATTGGA 816
Qy        666 AGAGAAAGCAGAATGAGAAACTTCAGGTTGTTAGGCATGAAGGAGATCCTGATTTTGAAG 725
Db        817 AGCAGAGACAAGAGAGGATGCACCAGACGGGGA---------ATGATGGTGGTGGTGATG 867
Qy        726 ATGGTGATGATGCTGATTTTCCAATGATGGATGAGGGAAGGCAGCCATTGTCTATGAAGA 785
Db        868 ATGGTGACGATGCTGATCTACCACTAATGGATGAAGCAAGACAACAACTGTCCAGGAAAA 927
Qy        786 TACCAATCAAATCGAGCAAGATAAATCCTTACCGGATGTTAATTGTGCTACGTCTTGTGA 845
Db        928 TTCCACTTCCATCAAGCCAGATTAATCCATATAGGATGATTATCATTATTCGGCTTGTGG 987
Qy        846 TTCTTGGTCTCTTCTTTCACTACCGTATTCTTCACCCCGTCAAAGATGCATATGCTTTGT 905
Db        988 TTTTGGGGTTCTTCTTCCACTACCGAGTGATGCATCCGGTGAATGATGCATTTGCTTTGT 1047
Qy        906 GGCTTATTTCTGTTATATGTGAGATATGGTTTGCTGTTTCATGGGTTCTTGATCAGTTCC 965
Db       1048 GGCTCATATCTGTTATCTGTGAAATCTGGTTTGCCATGTCTTGGATTCTTGATCAATTCC 1107
Qy        966 CTAAATGGTACCCTATCGAGCGAGAAACGTACTTGGACCGACTCTCATTAAGATATGAGA 1025
```



```
y       3174 TTTGGGTCCGGGTTAATCCGTTTGTGGCGA---AAGGCGGTCCTATTCTCGAGATCTGTG 3230
             ||||||| ||| || |||| ||  | ||||    | |  |||||  |||| |||   ||||
b       3328 TTTGGGTTCGGATTGATCCTTTCCTTGCGAAGGATGATGGTCCGCTTCTTGAGGAGTGTG 3387

)y      3231 GTTTAGACTGC 3241
             |||| || |||
)b      3388 GTTTGGATTGC 3398


ESULT 7
S-10-160-719-37
 Sequence 37, Application US/10160719
 Publication No. US20030167528A1
 GENERAL INFORMATION:
  APPLICANT: Kanwarpal S. Dhugga
  APPLICANT:  Timothy G. Helentjaris
  APPLICANT:  Benjamin A.  Bowen
  APPLICANT:  Xun Wang
  TITLE OF INVENTION: Maize Cellulose Synthases and Uses
  TITLE OF INVENTION:  Thereof
  FILE REFERENCE: 0864
  CURRENT APPLICATION NUMBER: US/10/160,719
  CURRENT FILING DATE:  2002-06-03
  PRIOR APPLICATION NUMBER: US/09/371,383A
  PRIOR FILING DATE: 1999-08-06
  PRIOR APPLICATION NUMBER: US 60/096,822
  PRIOR FILING DATE: 1998-08-17
  NUMBER OF SEQ ID NOS: 60
  SOFTWARE: FastSEQ for Windows Version 3.0
 SEQ ID NO 37
   LENGTH: 3969
   TYPE: DNA
   ORGANISM: Zea mays
   FEATURE:
   NAME/KEY: CDS
   LOCATION: (144)...(3401)
)S-10-160-719-37

 Query Match            43.3%;  Score 1489.8;  DB 12;  Length 3969;
 Best Local Similarity  67.3%;  Pred. No. 0;
 Matches 2189;  Conservative    0;  Mismatches 1032;  Indels   30;  Gaps    5;

)y         9 GGTTAATCGCCGGTTCTCACAACAGGAATGAGTTTGTCCTCATTAATGCCGATGAGAATG 68
             || |  | ||||| || |||||| | || ||| | ||| ||||      ||| |   ||
b        160 GGCTGGTGGCCGGCTCCCACAACCGCAACGAGCTCGTCGTCATCCGCCGCGACGGCGATC 219

)y        69 CCCGAATAAGATCAGTCCAAGAGCTGAGTGGACAGACATGTCAAATCTGCAGAGATGAGA 128
             || |    |   |     | |||| ||  || |||   || || || ||| | || ||
b        220 CCGGGCCGAAGCCGCCGCGGGAGCAGAACGGGCAGGTGTGCCAGATTTGCGGCGACGACG 279

)y       129 TCGAATTGACTGTTGATGGAGAACCGTTTGTGGCATGTAACGAATGTGCATTCCCTGTGT 188
             |||   |  |    | || || || || ||||| || ||||| || || ||||| || |
)b       280 TCGGCCTTGCCCCCGGCGGGGACCCCTTCGTGGCGTGCAACGAGTGCGCCTTCCCCGTCT 339

)y       189 GTAGACCTTGCTATGAGTACGAAAGACGAGAAGGCAATCAAGCTTGTCCACAGTGCAAAA 248
             | |   ||||| || |||||  | || || ||||  ||    || || ||||||||  |
)b       340 GCCGGGACTGCTACGAATACGAGCGCCGGGAGGGCACGCAGAACTGCCCCCAGTGCAAGA 399

)y       249 CCCGTTTCAAACGTCTTAAAGGAAGTCCAAGAGTTGAAGGTGATGAAGAGGAAGATGACA 308
             | || | ||| || || || ||  | | | | ||     ||||| || ||||| || | |
)b       400 CTCGATACAAGCGCCTCAAGGGCTGCCAACGTGTGACCGGTGACGAGGAGGAGGACGGCG 459

)y       309 TTGATGATTTAGACAATGAGTTTGA---GTATGGAAATAATGGGATTGGATTTGATCAGG 365
             | |||||  | ||||| |||||  |   | | ||  || |    |       | |   ||
)b       460 TCGATGACCTGGACAACGAGTTCAACTGGGACGGCCATGACTCGCAGTCTGTGGCCGAGT 519

)y       366 TTTCTGAAGGTATGTCAATCTCTCGTCGCAACTCCGGTTTCCCACAATCTGATTTGGATT 425
                          ||          ||      ||| ||| |   |       |
)b       520 CCATGCTCTACGGCCACATGAGCTACGGCCGTGGAGGTGACCCTAATGGCGCGCCACAAG 579

)y       426 CAGCTCCACCTGGCTCTCAGATTCCATTGCTGACTTACGGCGACGAGGACGTTGAGATTT 485
             |    | |    | | |  ||||| | || ||  |||  |   ||  | ||| ||
)b       580 CTTTCCAGCTCAACCCCAATGTTCCACTCCTCACCAACGGGCAAATGGTGGATGACATCC 639

Qy       486 CTTCTGATAGACATGCTCTTATTGTTCCTCCTTCACTTGGTGGTCATGGCAATAGAGTTC 545
             |  | ||    || || |  | || ||  ||| | ||||||   || || ||    | |
Db       640 CACCGGAGCAGCACGCGC---TGGTGCCTTCTTTCATGGGTGGTGGGGGAAAGAGGATAC 696

Qy       546 ATCCTGTTTCTCTTTCTGACCCGACCGTGGCTGCACATCGAAGGCTGATGGTACCTCAGA 605
             |||| || ||  | | || || | |  ||| || |  |||   |||| ||    |
Db       697 ATCCCCTTCCTTATGCGGATCCCAGCTTACCTGTGCAACCCAGGTCTATGGACCCATCCA 756

Qy       606 AAGATCTTGCGGTTTATGGTTATGGAAGTGTCGCTTGGAAAGATCGGATGGAGGAATGGA 665
             | |||||||| |  ||||| |||| |||||  |||||| || |||||||| | ||||||
Db       757 AGGATCTTGCTGCATATGGGTATGGTAGTGTTGCTTGGAAGGAACGGATGGAGAATTGGA 816

Qy       666 AGAGAAAGCAGAATGAGAAACTTCAGGTTGTTAGGCATGAAGGAGATCCTGATTTTGAAG 725
             ||   |  ||  |  ||  |  |||  |  |          |||  ||  :  ||| |
Db       817 AGCAGAGACAAGAGAGGATGCACCAGACGGGGA---------ATGATGGTGGTGGTGATG 867

Qy       726 ATGGTGATGATGCTGATTTTCCAATGATGGATGAGGGAAGGCAGCCATTGTCTATGAAGA 785
             |||||||  |||||| | | | | | |||| | ||| || | | |||| | ||| |
Db       868 ATGGTGACGATGCTGATCTACCACTAATGGATGAAGCAAGACAACAACTGTCCAGGAAAA 927

Qy       786 TACCAATCAAATCGAGCAAGATAAATCCTTACTGGATGTTAATTGTGCTACGTCTTGTGA 845
             | ||| |  ||  | ||| ||  | ||  |||||  | ||  |  | :  ||||
Db       928 TTCCACTTCCATCAAGCCAGATTAATCCATATAGGATGATTATCATTATTCGGCTTGTGG 987

Qy       846 TTCTTGGTCTCTTCTTTCACTACCGTATTCTTCACCCCGTCAAAGATGCATATGCTTTGT 905
             ||  | ||  |||||| ||||||||  |  | || || || || |||||| ||||| ||
Db       988 TTTTGGGGTTCTTCTTCCACTACCGAGTGATGCATCCGGTGAATGATGCATTTGCTTTGT 1047

Qy       906 GGCTTATTTCTGTTATATGTGAGATATGGTTTGCTGTTTCATGGGTTCTTGATCAGTTCC 965
             |||| || |||||||| ||||| || :|||||||  | || ||| ||||||||| ||||
Db      1048 GGCTCATATCTGTTATCTGTGAAATCTGGTTTGCCATGTCTTGGATTCTTGATCAATTCC 1107

Qy       966 CTAAATGGTACCCTATCGAGCGAGAAACGTACTTGGACCGACTCTCATTAAGATATGAGA 1025
             | || |||| ||||| ||| |||| || ||  | ||||| || ||| | || | | | |
Db      1108 CAAAGTGGTTCCCTATTGAGAGAGAGACTTACCTAGACCGGCTGTCACTGAGGTTCGACA 1167

Qy      1026 AAGAAGGGAAACCGTCGGGACTATCCCCTGTGGATGTATTTGTTAGTACAGTGGATCCAT 1085
             | |||||  | || ||  |||  | ||  | ||| | ||||| ||||| || |||||  |
Db      1168 AGGAAGGCCAGCCATCTCAACTTGCTCCAATTGATTTCTTTGTCAGTACGGTTGATCCCT 1227

Qy      1086 TGAAAGAGCCTCCGCTTATTACTGCAAATACTGTCTTGTCTATTCTTGCTGTTGATTATC 1145
             | || || ||||  |  | || ||||||||  | ||||| ||| | || ||||||||
Db      1228 TAAAGGAACCTCCTTTGGTCACAACAAATACTGTTCTATCTATCCTTTCGGTGGATTATC 1287

Qy      1146 CTGTCGATAAGGTTGCTTGTTACGTATCTGATGATGGTGCTGCTATGCTTACTTTCGAAG 1205
             |||| |||||||| ||| || || ||||||||||||||| ||||| || || ||||
Db      1288 CTGTTGATAAGGTTTCTTGCTATGTTTCTGATGATGGTGCTGCAATGCTAACGTTTGAAG 1347

Qy      1206 CTCTTTCTGAGACCGCTGAATTCGCAAGGAAATGGGTTCCTTTCTGCAAGAAATATTGTA 1265
             | | ||||| ||  |||||| |||| ||||||||||||||||||||     || ||
Db      1348 CATTATCTGAAACATCTGAATTTGCAAAGAAATGGGTTCCTTTCTGCAAACGGTACAATA 1407

Qy      1266 TTGAGCCTCGTGCTCCCGAATGGTATTTCTGCCATAAAATGGACTACTTGAAGAATAAAG 1325
             |||| ||||| |||| || ||||| |||    || || || ||||||||||  | | |
Db      1408 TTGAACCTCGCGCTCCAGAGTGGTACTTCCAACAGAAGATAGACTACTTGAAAGACAAGG 1467

Qy      1326 TTCATCCCGCATTTGTTAGGGAGCGGCGAGCCATGAAGAGAGATTATGAAGAATTCAAAG 1385
             |     |  ||||||||||||| || |||| |||||||||| ||||| |||||||| |
Db      1468 TGGCAGCAAACTTTGTTAGGGAGAGGAGAGCAATGAAGAGAGAGTATGAGGAATTCAAGG 1527

Qy      1386 TAAAGATCAATGCTTTAGTAGCAACAGCACAGAAAGTGCCTGAGGATGGTTGGACTATGC 1445
             | |  |||||||| ||||| || | ||| |||||||| |||| || || ||||| ||||
Db      1528 TGAGAATCAATGCCTTAGTTGCCAAAGCCCAGAAAGTTCCTGAAGAAGGATGGACAATGC 1587

Qy      1446 AAGACGGTACACCTTGGCCCGGTAATAGTGTGCGAGATCATCCTGGCATGATTCAGGTCT 1505
             |||| || || || |||| || || | ||| || |||||||||| ||||||||||||||
Db      1588 AAGATGGAACCCCCTGGCCTGGAAACAATGTTCGTGATCATCCTGGAATGATTCAGGTCT 1647

Qy      1506 TCCTTGGAAGTGACGGTGTTCGTGATGTCGAAAACAACGAGTTGCCTCGATTAGTTTACG 1565
             |||||||     ||| |  | |||    ||  || ||  |||| ||||| ||||| |
Db      1648 TCCTTGGCCAAAGCGGAGGCCTTGACTGTGAGGGAAATGAACTGCCACGATTGGTTTATG 1707
```

```
Qy       1566 TTTCTCGTGAGAAGAGACCCGGATTTGATCACCATAAGAAGGCTGGAGCTATGAATTCCC 1625
Db       1708 TTTCTAGAGAGAAACGACCAGGCTATAACCATCATAAGAAAGCTGGTGCTATGAATGCAT 1767

Qy       1626 TGATACGAGTCTCTGGGGTTCTATCAAATGCTCCTTACCTTCTGAATGTCGATTGTGATC 1685
Db       1768 TGGTCCGAGTCTCTGCTGTACTAACAAATGCTCCATATTTGTTAAACTTGGATTGTGATC 1827

Qy       1686 ACTACATCAACAATAGCAAAGCTCTTAGAGAAGCAATGTGTTTCATGATGGATCCTCAGT 1745
Db       1828 ACTACATCAACAACAGCAAGGCTATAAAGGAAGCAATGTGTTTTATGATGGACCCTTTAC 1887

Qy       1746 CAGGAAAGAAAATCTGTTATGTTCAGTTCCCTCAAAGGTTCGATGGGATTGATAGGCACG 1805
Db       1888 TAGGAAAGAAGGTTTGCTATGTACAGTTCCCTCAAAGATTTGATGGGATTGATCGCCATG 1947

Qy       1806 ATCGATACTCAAATCGCAATGTTGTGTTCTTTGATATCAATATGAAAGGTTTGGATGGGC 1865
Db       1948 ACCGATATGCTAACCGGAATGTTGTCTTTTTTGATATCAACATGAAAGGTTTGGATGGTA 2007

Qy       1866 TACAAGGGCCTATATACGTCGGTACAGGTTGTGTTTTCAGGAGGCAAGCGCTTTACGGAT 1925
Db       2008 TTCAGGGTCCAATTTATGTTGGTACTGGATGTGTATTTAGAAGGCAGGCATTATATGGTT 2067

Qy       1926 TTGATGCACCGAAGAAGAAGAAGGGCCCACGTAAGACATGCAATTGCTGGCCAAAATGGT 1985
Db       2068 ATGATGCCCCCAAAACAAAGAAGCCACCATCAAGGACTTGCAACTGCTGGCCCAAGTGGT 2127

Qy       1986 GTCTCCTATGTTTTGGTTCAAGAAAGAATCGTAAAGCAAAGACAGTGGCTGCGGATAAGA 2045
Db       2128 GCTTTTGCTGTTGCTGCTTTGGCAATAGGAAGCAAAAGAAGACTACCAAACCCAAAACAG 2187

Qy       2046 AGAAGAAGAA----------TAGGGAAGCGTCAAAGCAGATCCACGCATTAGAAAATA 2093
Db       2188 AGAAGAAAAAGTTATTATTTTTCAAGAAAGAAGAGAACCAATCCCCTGCATATGCTCTTG 2247

Qy       2094 TCGAAGAGGGCCGCGGTCATAAAGTTCTTAACGTAGAACAGTCAACCGAGGCAATGCAAA 2153
Db       2248 GTGAAATTGACGAAGCTGCTCCAGGAGCTGAGAATGAAAAGGCCGGTATTGTAAATCAAC 2307

Qy       2154 TGAAGTTGCAGAAGAAATATGGGCAGTCTCCTGTATTTGTTGCATCTGCGCGTCTGGAGA 2213
Db       2308 AAAAATTAGAAAAGAAATTTGGCCAATCTTCTGTTTTTGTTACATCCACACTTCTCGAGA 2367

Qy       2214 ATGGTGGGATGGCTAGAAACGCAAGCCCGGCTTGTCTGCTTAAAGAAGCCATCCAAGTCA 2273
Db       2368 ATGGTGGAACCTTGAAGAGTGCAAGTCCTGCTTCTCTTTTGAAAGAAGCTATACATGTCA 2427

Qy       2274 TTAGTCGCGGATATGAAGATAAAACTGAATGGGGAAAAGAGATTGGGTGGATCTATGGTT 2333
Db       2428 TTAGTTGTGGTTATGAAGACAAGACAGACTGGGGAAAAGAGATTGGCTGGATCTATGGAT 2487

Qy       2334 CTGTTACCGAAGATATTCTTACGGGTTCTAAGATGCATTCTCATGGTTGGAGACATGTTT 2393
Db       2488 CAGTTACAGAAGATATTCTAACTGGTTTCAAGATGCATTGTCATGGTTGGCGGTCAATTT 2547

Qy       2394 ATTGTACACCAAAGTTAGCGGCTTTCAAAGGATCAGCTCCAATCAATCTTTCGGATCGTC 2453
Db       2548 ACTGCATACCTAAACGGGTTGCATTCAAAGGTTCTGCACCTCTGAATCTTTCAGATCGTC 2607

Qy       2454 TCCATCAAGTTCTTCGATGGGCGCTTGGGTCGGTTGAGATTTTCTTGAGTAGGCATTGTC 2513
Db       2608 TTCACCAGGTGCTTCGGTGGGCTCTTGGGTCTATTGAGATCTTCTTCAGCAATCATTGCC 2667

Qy       2514 CTATTTGGTATGGTTATGGAGGTGGGTTGAAATGGCTTGAGCGGTTGTCCTACATTAACT 2573
Db       2668 CTCTTTGGTATGGGTATGGTGGCGGTCTGAAATTTTTGGAAAGATTTTCCTACATCAACT 2727

Qy       2574 CTGTGGTTTACCCGTGGACCTCTCTACGGCTCATCGTTTACTGTTCTCTCCCTGCCATCT 2633
Db       2728 CCATCGTGTATCCTTGGACATCTATTCCCCTCTTGGCTTACTGTACATTGCCTGCCATCT 2787

Qy       2634 GTCTTCTCACTGGAAAATTCATCGTTCCCGAGATTAGCAACTATGCGAGTATCCTCTTCA 2693
```

```
Db       2788 GTTTATTGACAGGGAAATTTATCACTCCAGAGCTGAATAATGTTGCCAGCCTGTGGTTCA 2847

Qy       2694 TGGCGCTCTTCTCGTCGATTGCAATAACGGGTATTCTCGAGATGCAATGGGGCAAAGTTG 2753
Db       2848 TGTCACTTTTTATCTGCATTTTTGCTACGAGCATCCTAGAAATGAGATGGAGTGGTGTTG 2907

Qy       2754 GGATCGATGATTGGTGGAGAAACGAACAGTTTTGGGTCATTGGAGGTGTTTCTGCGCATC 2813
Db       2908 GAATTGATGACTGGTGGAGGAATGAGCAGTTCTGGGTCATTGGAGGTGTGTCCTCACACC 2967

Qy       2814 TGTTTGCTCTCTTCCAAGGTCTCCTCAAGGTTCTTGCTGGTGTCGACACTAACTTCACAG 2873
Db       2968 TCTTTGCTGTGTTCCAGGGACTTCTCAAGGTCATAGCTGGTGTTGATACAAGCTTCACCG 3027

Qy       2874 TCACATCAAAAGCAGCTGATGATGGAGAGTTCTCTGACCTTTACCTCTTCAAATGGACTT 2933
Db       3028 TGACATCAAAGGGTGGAGATGATGAGGAGTTCTCAGAGCTATATACATTCAAATGGACTA 3087

Qy       2934 CACTTCTCATCCCTCCAATGACTCTACTCATCATAAACGTCATTGGAGTCATAGTCGGAG 2993
Db       3088 CCTTATTGATACCTCCTACCACCTTGCTTCTATTGAACTTCATTGGTGTGGTCGCTGGCG 3147

Qy       2994 TCTCTGATGCCATCAGCAATGGATACGACTCGTGGGGACCGCTTTTCGGAAGACTGTTCT 3053
Db       3148 TTTCAAATGCGATCAATAACGGATATGAGTCATGGGGCCCCCTCTTTGGGAAGCTATTCT 3207

Qy       3054 TTGCACTTTGGGTCATCATTCATCTTTACCCGTTCCTTAAAGGTTTGCTTGGGAAACAAG 3113
Db       3208 TTGCATTTTGGGTGATTGTCCATCTTTATCCCTTTCTCAAAGGTTTGGTTGGAAGGCAAA 3267

Qy       3114 ATAGAATGCCAACCATTATTGTCGTCTGGTCCATCCTCCTGGCCTCGATTCTTACACTTC 3173
Db       3268 ACAGGACACCAACGATTGTCATCGTCTGGTCCATTCTGCTGGCTTCAATCTTCTCGCTCC 3327

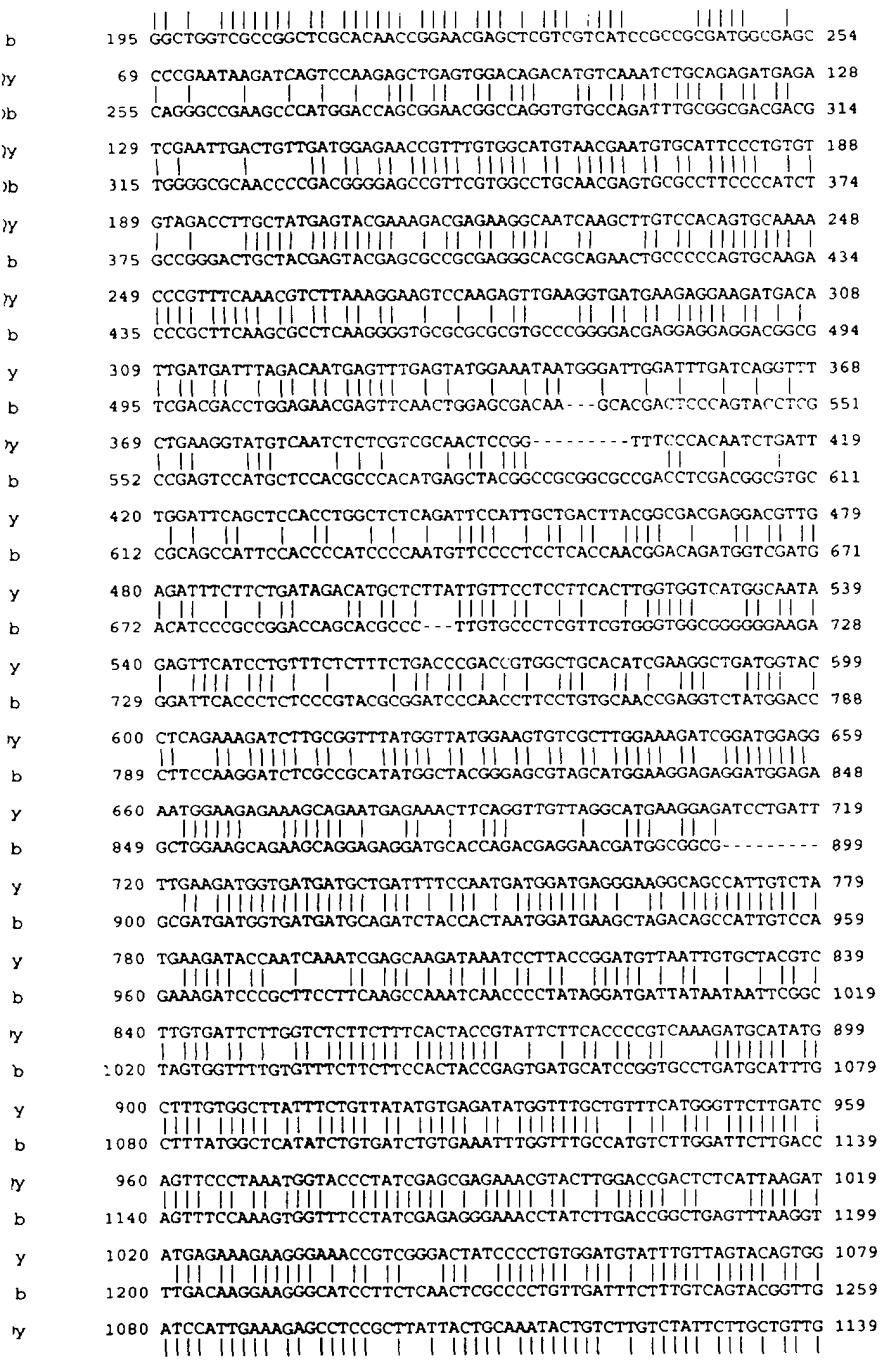
Qy       3174 TTTGGGTCCGGGTTAATCCGTTTGTGGCGA---AAGGCGGTCCTATTCTCGAGATCTGTG 3230
Db       3328 TTTGGGTTCGGATTGATCCTTTCCTTGCGAAGGATGATGGTCCGCTTCTTGAGGAGTGTG 3387

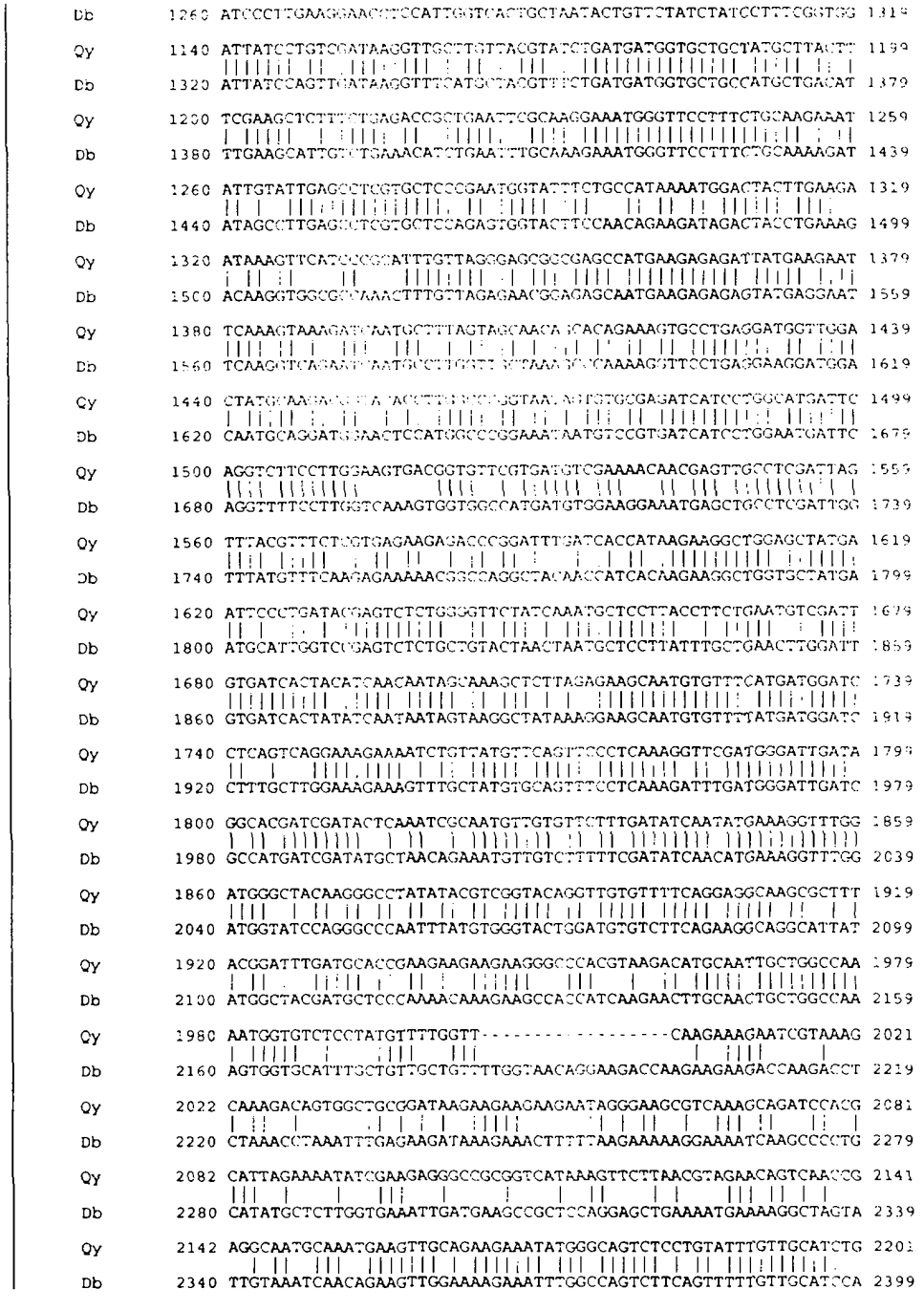
Qy       3231 GTTTAGACTGC 3241
Db       3388 GTTTGGATTGC 3398
```

```
RESULT 8
US-09-900-237-3
; Sequence 3, Application US/09900237
; Patent No. US20020120124A1
; GENERAL INFORMATION:
;  APPLICANT: Allen, Stephen
;  TITLE OF INVENTION: Plant Cellulose Synthases
;  FILE REFERENCE: BB1170 US CIP
;  CURRENT APPLICATION NUMBER: US/09/900,237
;  CURRENT FILING DATE:  2001-07-06
;  PRIOR APPLICATION NUMBER: 60/092,844
;  PRIOR FILING DATE: 1998-07-14
;  PRIOR APPLICATION NUMBER: PCT/US99/15871
;  PRIOR FILING DATE: 1999-07-13
;  PRIOR APPLICATION NUMBER: 09/720383
;  PRIOR FILING DATE: 2000-12-21
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 3
;   LENGTH: 3776
;   TYPE: DNA
;   ORGANISM: Zea mays
US-09-900-237-3

  Query Match           43.2%;  Score 1489;  DB 10;  Length 3776;
  Best Local Similarity   67.5%;  Pred. No. 0;
  Matches 2203;  Conservative    0;  Mismatches 1015;  Indels   45;  Gaps    6;

Qy          9 GGTTAATCGCCGGTTCTCACAACAGGAATGAGTTTGTCCTCATTAATGCCGATGAGAATG 68
```





```
Qy       2202 CGCGTCTGGAGAATGGTGGGATGGCTAGAAACGCAAGCCCGGCTTGTCTGCTTAAAGAAG 2261
Db       2400 CACTTCTTGAGAATGGTGGAACCCTGAAGAGTGCCAGTCCAGCTTCTCTTCTGAAGGAAG 2459

Qy       2262 CCATCCAAGTCATTAGTCGCGGATATGAAGATAAAACTGAATGGGGAAAAGAGATTGGGT 2321
Db       2460 CTATACATGTCATCAGTTGTGGATATGAAGACAAAACAGGCTGGGGAAAAGATATTGGTT 2519

Qy       2322 GGATCTATGGTTCTGTTACCGAAGATATTCTTACGGGTTCTAAGATGCATTCTCATGGTT 2381
Db       2520 GGATTTATGGATCAGTCACAGAAGATATTCTTACTGGGTTTAAGATGCACTGCCATGGTT 2579

Qy       2382 GGAGACATGTTTATTGTACACCAAAGTTAGCGGCTTTCAAAGGATCAGCTCCAATCAATC 2441
Db       2580 GGCGGTCAATTTACTGCATACCTAAACGGGCCGCCTTCAAAGGTTCCGCACCTCTCAATC 2639

Qy       2442 TTTCGGATCGTCTCCATCAAGTTCTTCGATGGGCGCTTGGGTCGGTTGAGATTTTCTTGA 2501
Db       2640 TTTCCGATCGTCTTCACCAGGTTCTTCGGTGGGCTCTTGGTTCAATTGAAATTTTCTTCA 2699

Qy       2502 GTAGGCATTGTCCTATTTGGTATGGTTATGGAGGTGGGTTGAAATGGCTTGAGCGGTTGT 2561
Db       2700 GCAACCACTGCCCTCTCTGGTATGGGTATGGTGGTGGACTAAAGTTCCTGGAAAGGTTTT 2759

Qy       2562 CCTACATTAACTCTGTGGTTTACCCGTGGACCTCTCTACCGCTCATCGTTTACTGTTCTC 2621
Db       2760 CGTACATTAACTCCATCGTATACCCTTGGACATCTATCCCGCTCTTGGCCTATTGCACAT 2819

Qy       2622 TCCCTGCCATCTGTCTTCTCACTGGAAAATTCATCGTTCCCGAGATTAGCAACTATGCGA 2681
Db       2820 TGCCTGCCATCTGCTTGCTGACAGGGAAATTTATCACGCCAGAGCTTAACAATGTTGCCA 2879

Qy       2682 GTATCCTCTTCATGGCGCTCTTCTCGTCGATTGCAATAACGGGTATTCTCGAGATGCAAT 2741
Db       2880 GCCTCTGGTTCATGTCACTTTTCATCTGCATTTTTGCTACGAGCATCCTGGAAATGAGAT 2939

Qy       2742 GGGGCAAAGTTGGGATCGATGATTGGTGGAGAAACGAACAGTTTTGGGTCATTGGAGGTG 2801
Db       2940 GGAGTGGTGTAGGCATCGATGACTGGTGGAGAAACGAGCAGTTTTGGGTCATTGGAGGCG 2999

Qy       2802 TTTCTGCGCATCTGTTTGCTCTCTTCCAAGGTCTCCTCAAGGTTCTTGCTGGTGTCGACA 2861
Db       3000 TGTCTTCACATCTCTTTGCTGTGTTCCAGGGACTCCTCAAGGTCATAGCTGGTGTAGACA 3059

Qy       2862 CTAACTTCACAGTCACATCAAAAGCAGCTGATGATGGAGAGTTCTCTGACCTTTACCTCT 2921
Db       3060 CGAGCTTCACTGTGACATCCAAGGGCGGAGACGACGAGGAGTTCTCAGAGCTGTACACAT 3119

Qy       2922 TCAAATGGACTTCACTTCTCATCCCTCCAATGACTCTACTCATCATAAACGTCATTGGAG 2981
Db       3120 TCAAATGGACGACCCTTCTGATACCTCCGACAACCCTGCTCCTACTGAACTTCATTGGAG 3179

Qy       2982 TCATAGTCGGAGTCTCTGATGCCATCAGCAATGGATACGACTCGTGGGGACCGCTTTTCG 3041
Db       3180 TGGTAGCTGGCATCTCCAATGCGATCAACAACGGATATGAATCATGGGGCCCCCTGTTCG 3239

Qy       3042 GAAGACTGTTCTTTGCACTTTGGGTCATCATTCATCTTTACCCGTTCCTTAAAGGTTTGC 3101
Db       3240 GGAAGCTCTTCTTTGCATTTTGGGTGATCGTCCATCTTTACCCGTTCCTCAAGGGTCTGG 3299

Qy       3102 TTGGGAAACAAGATAGAATGCCAACCATTATTGTCGTCTGGTCCATCCTCCTGGCCTCGA 3161
Db       3300 TTGGGAGGCAGAACAGGACGCCAACGATTGTCATTGTCTGGTCCATCCTCCTGGCTTCGA 3359

Qy       3162 TTCTTACACTTCTTTGGGTCCGGGTTAATCCGTTTGTGGCGA--AAGGCGGTCCTATTC 3218
Db       3360 TCTTCTCGCTGCTTTGGGTCCGGATCGACCCGTTCCTTGCGAAGGATGATGGTCCCCTGT 3419

Qy       3219 TCGAGATCTGTGGTTTAGACTGC 3241
Db       3420 TGGAGGAGTGTGGTCTGGATTGC 3442
```

```
RESULT 9
US-10-209-059-41
; Sequence 41, Application US/10209059
; Publication No. US20030163838A1
; GENERAL INFORMATION:
;  APPLICANT: Dhugga, Kanwarpal S.
;  APPLICANT:  Wang, Haiyin
;  TITLE OF INVENTION: Maize Cellulose Synthases and Uses
;  TITLE OF INVENTION:  Thereof
;  FILE REFERENCE: 0864R2
;  CURRENT APPLICATION NUMBER: US/10/209,059
;  CURRENT FILING DATE:  2002-07-31
;  PRIOR APPLICATION NUMBER: 60/096,822
;  PRIOR FILING DATE: 1998-08-17
;  PRIOR APPLICATION NUMBER: 09/371,383
;  PRIOR FILING DATE: 1999-08-06
;  PRIOR APPLICATION NUMBER: 09/550,483
;  PRIOR FILING DATE: 2000-04-14
;  NUMBER OF SEQ ID NOS: 52
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 41
;   LENGTH: 3568
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 3487
;   OTHER INFORMATION: n = A,T,C or G
US-10-209-059-41

  Query Match             42.5%;  Score 1464.8;  DB 12;  Length 3568;
  Best Local Similarity   67.3%;  Pred. No. 0;
  Matches 2177;  Conservative    0;  Mismatches 1012;  Indels   45;  Gaps    6;

Qy         38 GAGTTTGTCCTCATTAATGCCGATGAGAATGCCCGAATAAGATCAGTCCAAGAGCTGAGT 97
Db         18 GAGCTCGTCGTCATCCGCCGCGATGGCGAGCCAGGGCCGAAGCCCATGGACCAGCGGAAC 77

Qy         98 GGACAGACATGTCAAATCTGCAGAGATGAGATCGAATTGACTGTTGATGGAGAACCGTTT 157
Db         78 GGCCAGGTGTGCCAGATTTGCGGCGACGACGTGGGGCGCAACCCCGACGGGGAGCCTTTC 137

Qy        158 GTGGCATGTAACGAATGTGCATTCCCTGTGTGTAGACCTTGCTATGAGTACGAAAGACGA 217
Db        138 GTGGCCTGCAACGAGTGCGCCTTCCCCATCTGCCGGGACTGCTACGAGTACGAGCGCCGC 197

Qy        218 GAAGGCAATCAAGCTTGTCCACAGTGCAAAACCCGTTTCAAACGTCTTAAAGGAAGTCCA 277
Db        198 GAGGGCACGCAGAACTGCCCCCAGTGCAAGACCCGCTTCAAGCGCTTCAAGGGGTGCGCG 257

Qy        278 AGAGTTGAAGGTGATGAAGAGGAAGATGACATTGATGATTTAGACAATGAGTTTGAGTAT 337
Db        258 CGCGTGCCCGGGGACGAGGAGGAGGACGGCGTCGACGACCTGGAGAACGAGTTCAACTGG 317

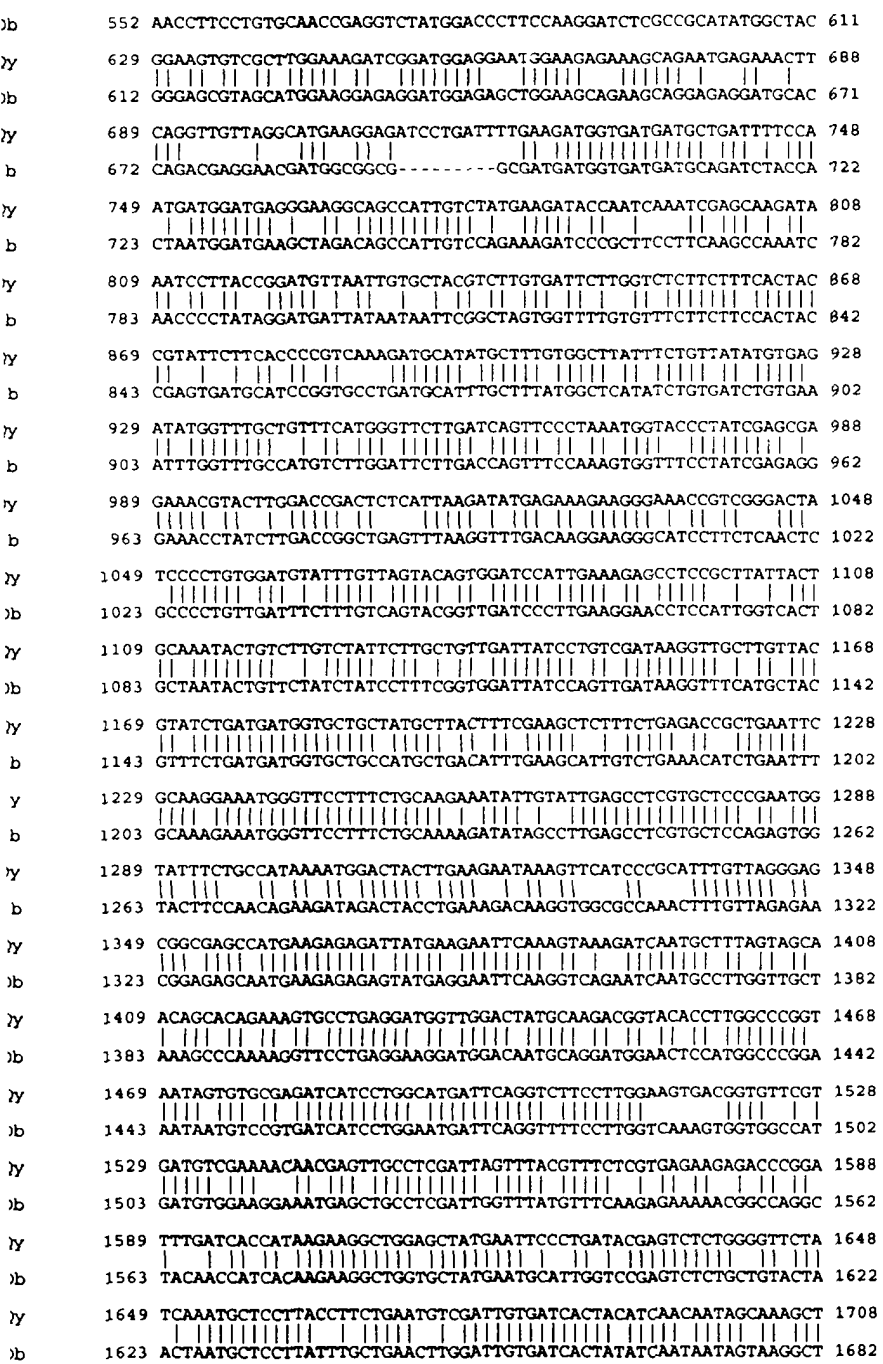
Qy        338 GGAAATAATGGGATTGGATTTGATCAGGTTTCTGAAGGTATGTCAATCTCTCGTCGCAAC 397
Db        318 AGCGACAA---GCACGACTCCCAGTACCTCGCCGAGTCCATGCTCCACGCCCACATGAGC 374

Qy        398 TCCGG---------TTTCCCACAATCTGATTTGGATTCAGCTCCACCTGGCTCTCAGATT 448
Db        375 TACGGCCGCGGCGCCGACCTCGACGGCGTGCCGCAGCCATTCCACCCCATCCCCAATGTT 434

Qy        449 CCATTGCTGACTTACGGCGACGAGGACGTTGAGATTTCTTCTGATAGACATGCTCTTATT 508
Db        435 CCCCTCCTCACCAACGGACAGATGGTCGATGACATCCCGCCGGACCAGCACGCCC---TT 491

Qy        509 GTTCCTCCTTCACTTGGTGGTCATGGCAATAGAGTTCATCCTGTTTCTCTTTCTGACCCG 568
Db        492 GTGCCCTCGTTCGTGGGTGGCGGGGGAAGAGGATTCACCCTCTCCCGTACGCGGATCCC 551

Qy        569 ACCGTGGCTGCACATCGAAGGCTGATGGTACCTCAGAAAGATCTTGCGGTTTATGGTTAT 628
```


Db 552 AACCTTCCTGTGCAACCGAGGTCTATGGACCCTTCCAAGGATCTCGCCGCATATGGCTAC 611
Qy 629 GGAAGTGTCGCTTGGAAAGATCGGATGGAGGAATGGAAGAGAAAGCAGAATGAGAAACTT 688
Db 612 GGGAGCGTAGCATGGAAGGAGAGGATGGAGAGCTGGAAGCAGAAGCAGGAGAGGATGCAC 671
Qy 689 CAGGTTGTTAGGCATGAAGGAGATCCTGATTTTGAAGATGGTGATGATGCTGATTTTCCA 748
Db 672 CAGACGAGGAACGATGGCGGCG---------GCGATGATGGTGATGATGCAGATCTACCA 722
Qy 749 ATGATGGATGAGGGAAGGCAGCCATTGTCTATGAAGATACCAATCAAATCGAGCAAGATA 808
Db 723 CTAATGGATGAAGCTAGACAGCCATTGTCCAGAAAGATCCCGCTTCCTTCAAGCCAAATC 782
Qy 809 AATCCTTACCGGATGTTAATTGTGCTACGTCTTGTGATTCTTGGTCTCTTCTTTCACTAC 868
Db 783 AACCCCTATAGGATGATTATAATAATTCGGCTAGTGGTTTTGTGTTTCTTCTTCCACTAC 842
Qy 869 CGTATTCTTCACCCCGTCAAAGATGCATATGCTTTGTGGCTTATTTCTGTTATATGTGAG 928
Db 843 CGAGTGATGCATCCGGTGCCTGATGCATTTGCTTTATGGCTCATATCTGTGATCTGTGAA 902
Qy 929 ATATGGTTTGCTGTTTCATGGGTTCTTGATCAGTTCCCTAAATGGTACCCTATCGAGCGA 988
Db 903 ATTTGGTTTGCCATGTCTTGGATTCTTGACCAGTTTCCAAAGTGGTTTCCTATCGAGAGG 962
Qy 989 GAAACGTACTTGGACCGACTCTCATTAAGATATGAGAAAGAAGGGAAACCGTCGGGACTA 1048
Db 963 GAAACCTATCTTGACCGGCTGAGTTTAAGGTTTGACAAGGAAGGGCATCCTTCTCAACTC 1022
Qy 1049 TCCCCTGTGGATGTATTTGTTAGTACAGTGGATCCATTGAAAGAGCCTCCGCTTATTACT 1108
Db 1023 GCCCCTGTTGATTTCTTTGTCAGTACGGTTGATCCCTTGAAGGAACCTCCATTGGTCACT 1082
Qy 1109 GCAAATACTGTCTTGTCTATTCTTGCTGTTGATTATCCTGTCGATAAGGTTGCTTGTTAC 1168
Db 1083 GCTAATACTGTTCTATCTATCCTTTCGGTGGATTATCCAGTTGATAAGGTTTCATGCTAC 1142
Qy 1169 GTATCTGATGATGGTGCTGCTATGCTTACTTTCGAAGCTCTTTCTGAGACCGCTGAATTC 1228
Db 1143 GTTTCTGATGATGGTGCTGCCATGCTGACATTTGAAGCATTGTCTGAAACATCTGAATTT 1202
Qy 1229 GCAAGGAAATGGGTTCCTTTCTGCAAGAAATATTGTATTGAGCCTCGTGCTCCCGAATGG 1288
Db 1203 GCAAAGAAATGGGTTCCTTTCTGCAAAAGATATAGCCTTGAGCCTCGTGCTCCAGAGTGG 1262
Qy 1289 TATTTCTGCCATAAAATGGACTACTTGAAGAATAAAGTTCATCCCGCATTTGTTAGGGAG 1348
Db 1263 TACTTCCAACAGAAGATAGACTACCTGAAAGACAAGGTGGCGCCAAACTTTGTTAGAGAA 1322
Qy 1349 CGGCGAGCCATGAAGAGAGATTATGAAGAATTCAAAGTAAAGATCAATGCTTTAGTAGCA 1408
Db 1323 CGGAGAGCAATGAAGAGAGAGTATGAGGAATTCAAGGTCAGAATCAATGCCTTGGTTGCT 1382
Qy 1409 ACAGCACAGAAAGTGCCTGAGGATGGTTGGACTATGCAAGACGGTACACCTTGGCCCGGT 1468
Db 1383 AAAGCCCAAAAGGTTCCTGAGGAAGGATGGACAATGCAGGATGGAACTCCATGGCCCGGA 1442
Qy 1469 AATAGTGTGCGAGATCATCCTGGCATGATTCAGGTCTTCCTTGGAAGTGACGGTGTTCGT 1528
Db 1443 AATAATGTCCGTGATCATCCTGGAATGATTCAGGTTTTCCTTGGTCAAAGTGGTGGCCAT 1502
Qy 1529 GATGTCGAAAACAACGAGTTGCCTCGATTAGTTTACGTTTCTCGTGAGAAGAGACCCGGA 1588
Db 1503 GATGTGGAAGGAAATGAGCTGCCTCGATTGGTTTATGTTTCAAGAGAAAAACGGCCAGGC 1562
Qy 1589 TTTGATCACCATAAGAAGGCTGGAGCTATGAATTCCCTGATACGAGTCTCTGGGGTTCTA 1648
Db 1563 TACAACCATCACAAGAAGGCTGGTGCTATGAATGCATTGGTCCGAGTCTCTGCTGTACTA 1622
Qy 1649 TCAAATGCTCCTTACCTTCTGAATGTCGATTGTGATCACTACATCAACAATAGCAAAGCT 1708
Db 1623 ACTAATGCTCCTTATTTGCTGAACTTGGATTGTGATCACTATATCAATAATAGTAAGGCT 1682
Qy 1709 CTTAGAGAAGCAATGTGTTTCATGATGGATCCTCAGTCAGGAAAGAAAATCTGTTATGTT 1768
Db 1683 ATAAAGGAAGCAATGTGTTTTATGATGGATCCTTTGCTTGGAAAGAAAGTTTGCTATGTG 1742
Qy 1769 CAGTTCCCTCAAAGGTTCGATGGGATTGATAGGCACGATCGATACTCAAATCGCAATGTT 1828
Db 1743 CAGTTTCCTCAAAGATTTGATGGGATTGATCGCCATGATCGATATGCTAACAGAAATGTT 1802
Qy 1829 GTGTTCTTTGATATCAATATGAAAGGTTTGGATGGGCTACAAGGGCCTATATACGTCGGT 1888
Db 1803 GTCTTTTTCGATATCAACATGAAAGGTTTGGATGGTATCCAGGGCCCAATTTATGTGGGT 1862
Qy 1889 ACAGGTTGTGTTTTCAGGAGGCAAGCGCTTTACGGATTTGATGCACCGAAGAAGAAGAAG 1948
Db 1863 ACTGGATGTGTCTTCAGAAGGCAGGCATTATATGGCTACGATGCTCCCAAAACAAAGAAG 1922
Qy 1949 GGCCCACGTAAGACATGCAATTGCTGGCCAAAATGGTGTCTCCTATGTTTTGGTT-- -- 2003
Db 1923 CCACCATCAAGAACTTGCAACTGCTGGCCAAAGTGGTGCATTTGCTGTTGCTGTTTGGT 1982
Qy 2004 -- --- CAAGAAAGAATCGTAAAGCAAAGACAGTGGCTGCGGATAAGAAGAAG 2050
Db 1983 AACAGGAAGACCAAGAAGAAGACCAAGACCTCTAAACCTAAATTTGAGAAGATAAAGAAA 2042
Qy 2051 AAGAATAGGGAAGCGTCAAAGCAGATCCACGCATTAGAAAATATCGAAGAGGGCCGCGGT 2110
Db 2043 CTTTTTAAGAAAAAGGAAAATCAAGCCCCTGCATATGCTCTTGGTGAAATTGATGAAGCC 2102
Qy 2111 CATAAAGTTCTTAACGTAGAACAGTCAACCGAGGCAATGCAAATGAAGTTGCAGAAGAAA 2170
Db 2103 GCTCCAGGAGCTGAAAATGAAAAGGCTAGTATTGTAAATCAACAGAAGTTGGAAAAGAAA 2162
Qy 2171 TATGGGCAGTCTCCTGTATTTGTTGCATCTGCGCGTCTGGAGAATGGTGGGATGGCTAGA 2230
Db 2163 TTTGGCCAGTCTTCAGTTTTTGTTGCATCCACACTTCTTGAGAATGGTGGAACCCTGAAG 2222
Qy 2231 AACGCAAGCCCGGCTTGTCTGCTTAAAGAAGCCATCCAAGTCATTAGTCGCGGATATGAA 2290
Db 2223 AGTGCCAGTCCAGCTTCTCTTCTGAAGGAAGCTATACATGTCATCAGTTGTGGATATGAA 2282
Qy 2291 GATAAAACTGAATGGGGAAAAGAGATTGGGTGGATCTATGGTTCTGTTACCGAAGATATT 2350
Db 2283 GACAAAACAGGCTGGGGAAAAGATATTGGTTGGATTTATGGATCAGTCACAGAAGATATT 2342
Qy 2351 CTTACGGGTTCTAAGATGCATTCTCATGGTTGGAGACATGTTTATTGTACACCAAAGTTA 2410
Db 2343 CTTACTGGGTTTAAGATGCACTGCCATGGTTGGCGGTCAATTTACTGCATACCTAAACGG 2402
Qy 2411 GCGGCTTTCAAAGGATCAGCTCCAATCAATCTTTCGGATCGTCTCCATCAAGTTCTTCGA 2470
Db 2403 GCCGCCTTCAAAGGTTCCGCACCTCTCAATCTTTCCGATCGTTTTCACCAGGTTCTTCGG 2462
Qy 2471 TGGGCGCTTGGGTCGGTTGAGATTTTCTTGAGTAGGCATTGTCCTATTTGGTATGGTTAT 2530
Db 2463 TGGGCTCTTGGTTCAATTGAAATTTTGTTCAGCAACCACTGCCCTCTCTGGTATGGGTAT 2522
Qy 2531 GGAGGTGGGTTGAAATGGCTTGAGCGGTTGTCCTACATTAACTCTGTGGTTTACCCGTGG 2590
Db 2523 GGTGGTGGACTAAAGTTCCTGGAAAGGTTTTCGTACATTAACTCCATCGTATACCCTTGG 2582
Qy 2591 ACCTCTCTACCGCTCATCGTTTACTGTTCTCTCCCTGCCATCTGTCTTCTCACTGGAAAA 2650
Db 2583 ACATCTATCCCGCTCTTGGCCTATTGCACATTGCCTGCCATCTGCTTGCTGACAGGGAAA 2642
Qy 2651 TTCATCGTTCCCGAGATTAGCAACTATGCGAGTATCCTCTTCATGGCGCTCTTCTCGTCG 2710
Db 2643 TTTATCACGCCAGAGCTTAACAATGTTGCCAGCCTCTGGTTCATGTCACTTTTCATCTGC 2702
Qy 2711 ATTGCAATAACGGGTATTCTCGAGATGCAATGGGGCAAAGTTGGGATCGATGATTGGTGG 2770
Db 2703 ATTTTTGCTACGAGCATCCTGGAAATGAGATGGAGTGGTGTAGGCATCGATGACTGGTGG 2762

```
Qy       2771 AGAAACGAACAGTTTTGGGTCATTGGAGGTGTTTCTGCGCATCTGTTTGCTCTCTTCCAA 2830
             |||||||| |||||||||||||||||||||| || ||| | ||||| |||||| | |||||
Db       2763 AGAAACGAGCAGTTTTGGGTCATTGGAGGCGTGTCTTCACATCTCTTTGCTGTGTTCCAG 2822

Qy       2831 GGTCTCCTCAAGGTTCTTGCTGGTGTCGACACTAACTTCACAGTCACATCAAAAGCAGCT 2890
             || |||||||||||  | |||||||| ||||| ! |||||| || ||||| || |  |
Db       2823 GGACTCCTCAAGGTCATAGCTGGTGTAGACACGAGCTTCACTGTGACATCCAAGGGCGGA 2882

Qy       2891 GATGATGGAGAGTTCTCTGACCTTTACCTCTTCAAATGGACTTCACTTCTCATCCCTCCA 2950
             || || |  |||||||| || || |||   ||||||||||  | |||||| || |||||
Db       2883 GACGACGAGGAGTTCTCAGAGCTGTACACATTCAAATGGACGACCCTTCTGATACCTCCG 2942

Qy       2951 ATGACTCTACTCATCATAAACGTCATTGGAGTCATAGTCGGAGTCTCTGATGCCATCAGC 3010
             |  || || ||| |  | ||| ||||||||||  |||  ||  ||||  |||| |||| |
Db       2943 ACAACCCTGCTCCTACTGAACTTCATTGGAGTGGTAGCTGGCATCTCCAATGCGATCAAC 3002

Qy       3011 AATGGATACGACTCGTGGGGACCGCTTTTCGGAAGACTGTTCTTTGCACTTTGGGTCATC 3070
             || ||||| || || ||||| || || ||||| |  || |||||||| ||||||| |||
Db       3003 AACGGATATGAATCATGGGGCCCCCTGTTCGGGAAGCTCTTCTTTGCATTTTGGGTGATC 3062

Qy       3071 ATTCATCTTTACCCGTTCCTTAAAGGTTTGCTTGGGAAACAAGATAGAATGCCAACCATT 3130
             | ||||||||||||||||| || ||| || |||||| ||  | || | |||||| |||
Db       3063 GTCCATCTTTACCCGTTCCTCAAGGGTCTGGTTGGGAGGCAGAACAGGACGCCAACGATT 3122

Qy       3131 ATTGTCGTCTGGTCCATCCTCCTGGCCTCGATTCTTACACTTCTTTGGGTCCGGGTTAAT 3190
             |  | |||||||||||||||||||| |||||  |  | || ||||||||||| |  |
Db       3123 GTCATTGTCTGGTCCATCCTCCTGGCTTCGATCTTCTCGCTGCTTTGGGTCCGGATCGAC 3182

Qy       3191 CCGTTTGTGGCGA---AAGGCGGTCCTATTCTCGAGATCTGTGGTTTAGACTGC 3241
             |||||  | ||||   | |  ||||  |  | |||   |||||| | || |||
Db       3183 CCGTTCCTTGCGAAGGATGATGGTCCCCTGTTGGAGGAGTGTGGTCTGGATTGC 3236


RESULT 10
US-10-160-719-1
; Sequence 1, Application US/10160719
; Publication No. US20030167528A1
; GENERAL INFORMATION:
;  APPLICANT: Kanwarpal S. Dhugga
;  APPLICANT:  Timothy G. Helentjaris
;  APPLICANT:  Benjamin A.  Bowen
;  APPLICANT:  Xun Wang
;  TITLE OF INVENTION: Maize Cellulose Synthases and Uses
;  TITLE OF INVENTION:  Thereof
;  FILE REFERENCE: 0864
;  CURRENT APPLICATION NUMBER: US/10/160,719
;  CURRENT FILING DATE:  2002-06-03
;  PRIOR APPLICATION NUMBER: US/09/371,383A
;  PRIOR FILING DATE: 1999-08-06
;  PRIOR APPLICATION NUMBER: US 60/096,822
;  PRIOR FILING DATE: 1998-08-17
;  NUMBER OF SEQ ID NOS: 60
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 3568
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (63)...(3239)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (3487)...(3487)
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(3568)
;   OTHER INFORMATION: n = A,T,C or G
US-10-160-719-1

  Query Match             42.5%;  Score 1464.8;  DB 12;  Length 3568;
  Best Local Similarity   67.3%;  Pred. No. 0;
  Matches 2177;  Conservative    0;  Mismatches 1012;  Indels   45;  Gaps    6;
```

```
Qy         38 GAGTTTGTCCTCATTAATGCCGATGAGAATGCCCGAATAAGATCAGTCCAAGAGCTGAGT 97
              ||| | ||| ||||       |||||  |  |  |  |  |  |  |||  ||
Db         18 GAGCTCGTCGTCATCCGCCGCGATGGCGAGCCAGGGCCGAAGCCCATGGACCAGCGGAAC 77

Qy         98 GGACAGACATGTCAAATCTGCAGAGATGAGATCGAATTGACTGTTGATGGAGAACCGTTT 157
              || |||   || || || ||| | || ||  | |      |    || || || || ||
Db         78 GGCCAGGTGTGCCAGATTTGCGGCGACGACGTGGGGCGCAACCCCGACGGGGAGCCTTTC 137

Qy        158 GTGGCATGTAACGAATGTGCATTCCCTGTGTGTAGACCTTGCTATGAGTACGAAAGACGA 217
              ||||| || |||||  || || |||||  | ||  |    ||||| |||||||  | ||
Db        138 GTGGCCTGCAACGAGTGCGCCTTCCCCATCTGCCGGGACTGCTACGAGTACGAGCGCCGC 197

Qy        218 GAAGGCAATCAAGCTTGTCCACAGTGCAAAACCCGTTTCAAACGTCTTAAAGGAAGTCCA 277
              || ||||  ||    || || |||||||| ||||| ||||| ||  | || ||  |  |
Db        198 GAGGGCACGCAGAACTGCCCCCAGTGCAAGACCCGCTTCAAGCGCTTCAAGGGGTGCGCG 257

Qy        278 AGAGTTGAAGGTGATGAAGAGGAAGATGACATTGATGATTTAGACAATGAGTTTGAGTAT 337
              | ||    || || || ||||| || | | | || ||  | || || |||||  | |
Db        258 CGCGTGCCCGGGGACGAGGAGGAGGACGGCGTCGACGACCTGGAGAACGAGTTCAACTGG 317

Qy        338 GGAAATAATGGGATTGGATTTGATCAGGTTTCTGAAGGTATGTCAATCTCTCGTCGCAAC 397
              |  | ||  |  |  |  |  |  |  |  |  |||  | | |    | |  |
Db        318 AGCGACAA---GCACGACTCCCAGTACCTCGCCGAGTCCATGCTCCACGCCCACATGAGC 374

Qy        398 TCCGG---------TTTCCCACAATCTGATTTGGATTCAGCTCCACCTGGCTCTCAGATT 448
              | |||          ||  |  |  | | ||  | ||  | | |  ||
Db        375 TACGGCCGCGGCGCCGACCTCGACGGCGTGCCGCAGCCATTCCACCCCATCCCCAATGTT 434

Qy        449 CCATTGCTGACTTACGGCGACGAGGACGTTGAGATTTCTTCTGATAGACATGCTCTTATT 508
              ||  | || ||  |||| |  || || ||| || |  | ||    || || |  ||
Db        435 CCCCTCCTCACCAACGGACAGATGGTCGATGACATCCCGCCGGACCAGCACGCCC---TT 491

Qy        509 GTTCCTCCTTCACTTGGTGGTCATGGCAATAGAGTTCATCCTGTTTCTCTTTCTGACCCG 568
              || ||  | |  | ||||  || || ||  |||| ||| |  |    | || ||
Db        492 GTGCCCTCGTTCGTGGGTGGCGGGGGGAAGAGGATTCACCCTCTCCCGTACGCGGATCCC 551

Qy        569 ACCGTGGCTGCACATCGAAGGCTGATGGTACCTCAGAAAGATCTTGCGGTTTATGGTTAT 628
              | | |  |||  || |  |||   ||||  |||   || ||||| || |  ||||| ||
Db        552 AACCTTCCTGTGCAACCGAGGTCTATGGACCCTTCCAAGGATCTCGCCGCATATGGCTAC 611

Qy        629 GGAAGTGTCGCTTGGAAAGATCGGATGGAGGAATGGAAGAGAAAGCAGAATGAGAAACTT 688
              || || || || ||||| ||  ||||||||   ||||||  ||||||  |   ||  |
Db        612 GGGAGCGTAGCATGGAAGGAGAGGATGGAGAGCTGGAAGCAGAAGCAGGAGAGGATGCAC 671

Qy        689 CAGGTTGTTAGGCATGAAGGAGATCCTGATTTTGAAGATGGTGATGATGCTGATTTTCCA 748
              |||      |  |||  || |          || ||||||||||||| ||| | |||
Db        672 CAGACGAGGAACGATGGCGGCG---------GCGATGATGGTGATGATGCAGATCTACCA 722

Qy        749 ATGATGGATGAGGGAAGGCAGCCATTGTCTATGAAGATACCAATCAAATCGAGCAAGATA 808
              | |||||||| |  || ||||||||||| |  ||||| ||  |    || ||| | ||
Db        723 CTAATGGATGAAGCTAGACAGCCATTGTCCAGAAAGATCCCGCTTCCTTCAAGCCAAATC 782

Qy        809 AATCCTTACCGGATGTTAATTGTGCTACGTCTTGTGATTCTTGGTCTCTTCTTTCACTAC 868
              || ||  |||||  | ||  |  | || || ||| || |  || ||||||| ||||||
Db        783 AACCCCTATAGGATGATTATAATAATTCGGCTAGTGGTTTTGTGTTTCTTCTTCCACTAC 842

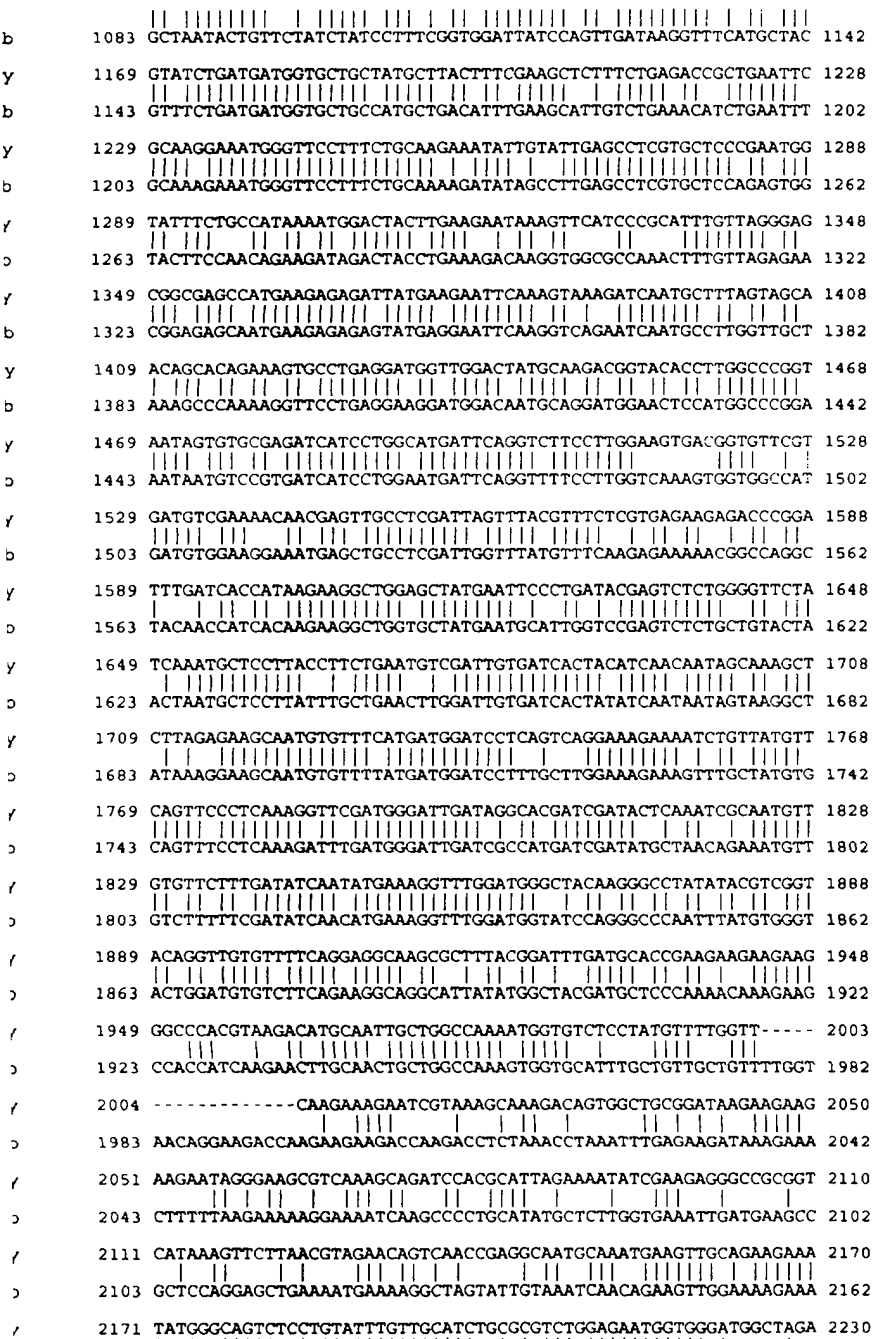
Qy        869 CGTATTCTTCACCCCGTCAAAGATGCATATGCTTTGTGGCTTATTTCTGTTATATGTGAG 928
              ||  |  | ||  ||  ||     |||||| ||||| ||||| ||  ||||| || |||||
Db        843 CGAGTGATGCATCCGGTGCCTGATGCATTTGCTTTATGGCTCATATCTGTGATCTGTGAA 902

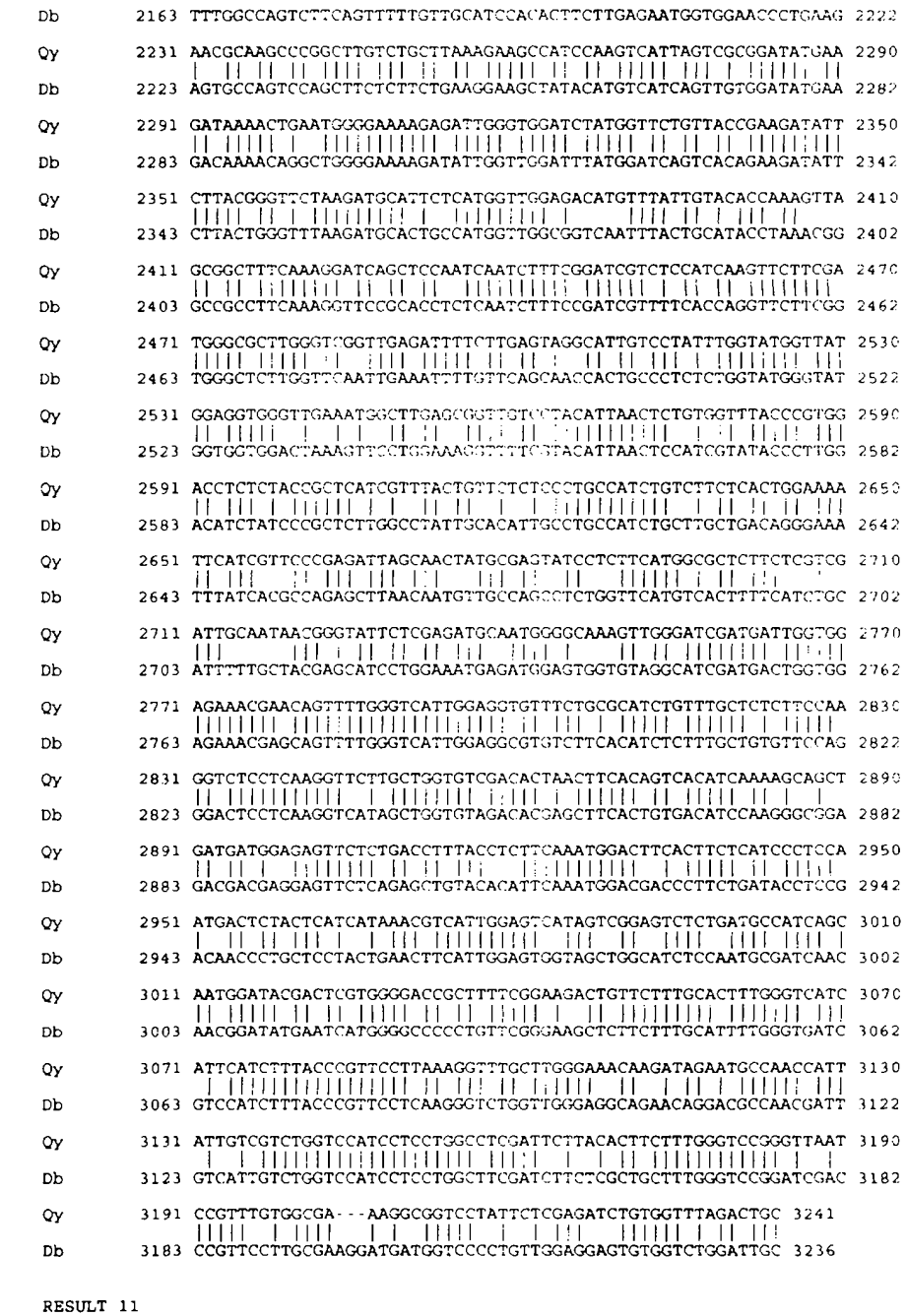
Qy        929 ATATGGTTTGCTGTTTCATGGGTTCTTGATCAGTTCCCTAAATGGTACCCTATCGAGCGA 988
              || ||||||||  | || ||| |||||||  ||||| || || ||||  |||||||||  |
Db        903 ATTTGGTTTGCCATGTCTTGGATTCTTGACCAGTTTCCAAAGTGGTTTCCTATCGAGAGG 962

Qy        989 GAAACGTACTTGGACCGACTCTCATTAAGATATGAGAAAGAAGGGAAACCGTCGGGACTA 1048
              ||||| ||  | ||||| ||     |||||  | ||| || ||||||  | || ||    |||
Db        963 GAAACCTATCTTGACCGGCTGAGTTTAAGGTTTGACAAGGAAGGGCATCCTTCTCAACTC 1022

Qy       1049 TCCCCTGTGGATGTATTTGTTAGTACAGTGGATCCATTGAAAGAGCCTCCGCTTATTACT 1108
              |||||||  ||| | |||||  |||||  ||  |||||  |||||  ||  |||||   |  | |||
Db       1023 GCCCCTGTTGATTTCTTTGTCAGTACGGTTGATCCCTTGAAGGAACCTCCATTGGTCACT 1082

Qy       1109 GCAAATACTGTCTTGTCTATTCTTGCTGTTGATTATCCTGTCGATAAGGTTGCTTGTTAC 1168
```

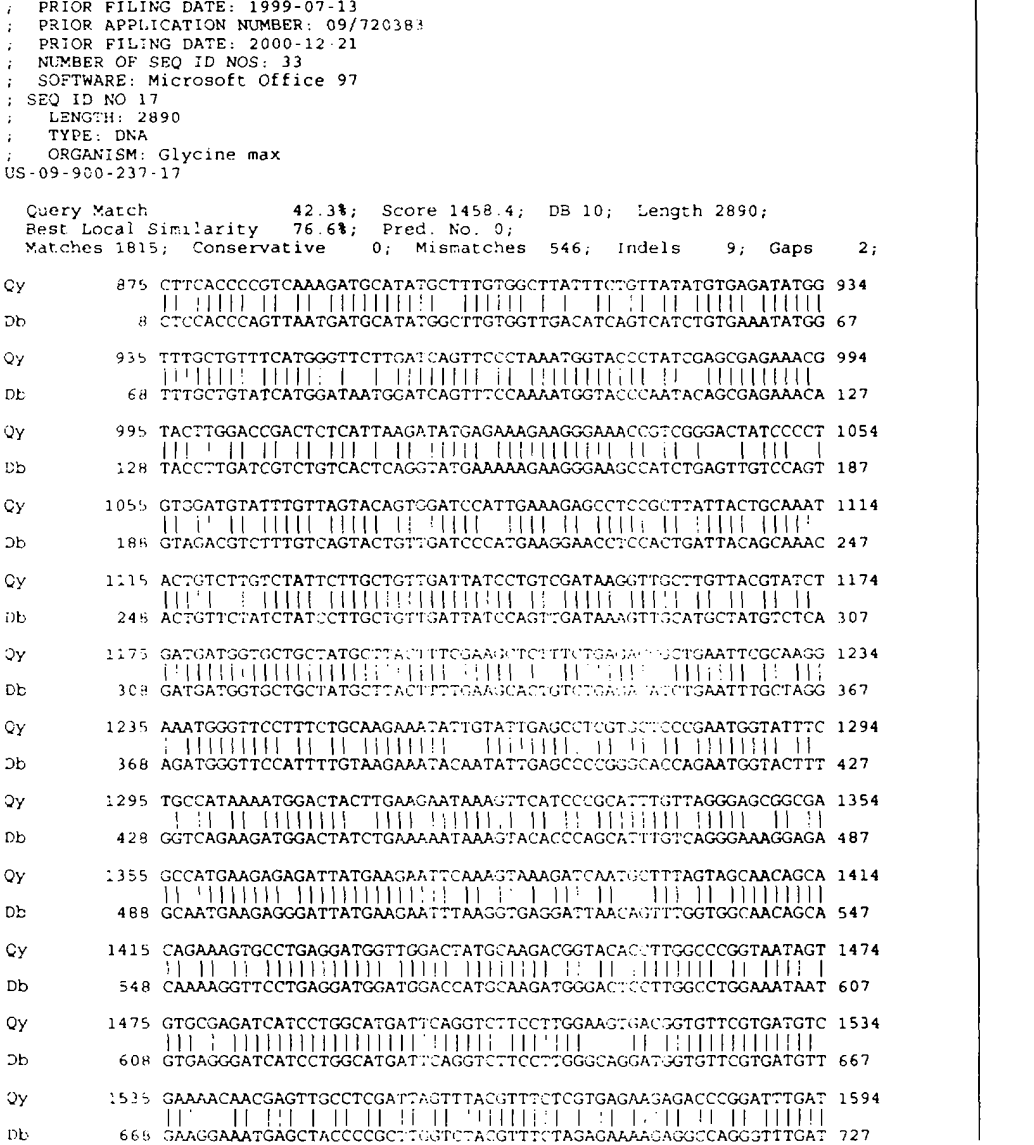


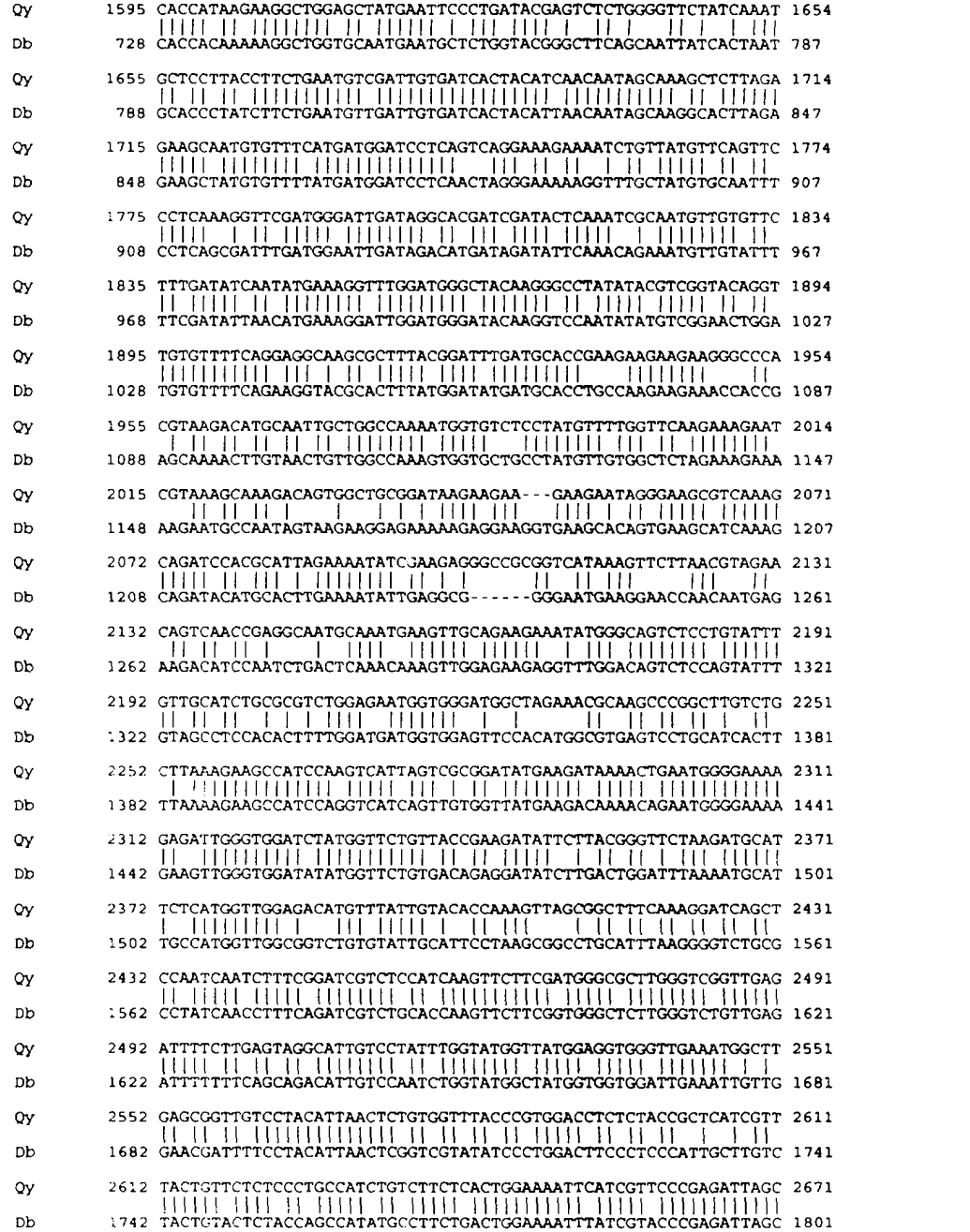


RESULT 11
US-09-900-237-17

```
; Sequence 17, Application US/09900237
; Patent No. US20020120124A1
; GENERAL INFORMATION:
;  APPLICANT: Allen, Stephen
;  TITLE OF INVENTION: Plant Cellulose Synthases
;  FILE REFERENCE: BB1170 US CIP
;  CURRENT APPLICATION NUMBER: US/09/900,237
;  CURRENT FILING DATE:  2001-07-06
;  PRIOR APPLICATION NUMBER: 60/092,844
;  PRIOR FILING DATE: 1998-07-14
;  PRIOR APPLICATION NUMBER: PCT/US99/15871
;  PRIOR FILING DATE: 1999-07-13
;  PRIOR APPLICATION NUMBER: 09/720383
;  PRIOR FILING DATE: 2000-12-21
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 17
;   LENGTH: 2890
;   TYPE: DNA
;   ORGANISM: Glycine max
US-09-900-237-17

  Query Match             42.3%;  Score 1458.4;  DB 10;  Length 2890;
  Best Local Similarity   76.6%;  Pred. No. 0;
  Matches 1815;  Conservative    0;  Mismatches  546;  Indels    9;  Gaps    2;
```





```
y       2672 AACTATGCGAGTATCCTCTTCATGGCGCTCTTCTCGTCGATTGCAATAACGGGTATTCTC 2731
b       1802 AACTATGCCAGTCTTGTGTTCATGGCCCTCTTCATATCCATTGCAGCAACTGGCATCCTT 1861

y       2732 GAGATGCAATGGGGCAAAGTTGGGATCGATGATTGGTGGAGAAACGAACAGTTTTGGGTC 2791
b       1862 GAGATGCAATGGGGCGGTGTTAGCATAGACGACTGGTGGAGGAACGAACAGTTTTGGGTG 1921

y       2792 ATTGGAGGTGTTTCTGCGCATCTGTTTGCTCTCTTCCAAGGTCTCCTCAAGGTTCTTGCT 2851
b       1922 ATCGGAGGTGTTTCTTCCCATCTATTTGCCCTATTTCAGGGTTTACTGAAGGTCTTGGCT 1981

y       2852 GGTGTCGACACTAACTTCACAGTCACATCAAAAGCAGCTGATGATGGAGACTTCTCTGAC 2911
b       1982 GGTGTGAACACAAACTTCACTGTGACCTCAAAAGCAGCAGATGATGGAGAATTCTCAGAA 2041

y       2912 CTTTACCTCTTCAAATGGACTTCACTTCTCATCCCTCCAATGACTCTACTCATCATAAAC 2971
b       2042 CTCTACATATTCAAGTGGACATCACTCTTGATCCCTCCAATGACTTTACTTATCATGAAT 2101

y       2972 GTCATTGGAGTCATAGTCGGAGTCTCTGATGCCATCAGCAATGGATACGACTCGTGGGGA 3031
b       2102 ATTGTCGGGGTGGTTGTCGGGATCTCAGATGCCATCAACAATGGTTATGACTCATGGGGA 2161

y       3032 CCGCTTTTCGGAAGACTGTTCTTTGCACTTTGGGTCATCATTCATCTTTACCCGTTCCTT 3091
b       2162 CCTCTCTTTGGTAGATTGTTCTTTGCATTGTGGGTGATCCTCCATCTCTACCCCTTCTTG 2221

y       3092 AAAGGTTTGCTTGGGAAACAAGATAGAATGCCAACCATTATTGTCGTCTGGTCCATCCTC 3151
b       2222 AAGGGGTTGCTTGGAAAACAAGATAGAATGCCAACCATTATATTGGTTTGGTCAATCCTT 2281

y       3152 CTGGCCTCGATTCTTACACTTCTTTGGGTCCGGGTTAATCCGTTTGTGGCGAAAGGCGGT 3211
b       2282 CTGGCCTCCATCTTGACTCTCATGTGGGTCAGAATTAACCCGTTTGTGTCAAGAGACGGC 2341

y       3212 CCTATTCTCGAGATCTGTGGTTTAGACTGC 3241
b       2342 CCCGTGTTAGAAATTTGTGGATTGAATTGC 2371


ESULT 12
S-09-900-237-7
 Sequence 7, Application US/09900237
 Patent No. US20020120124A1
 GENERAL INFORMATION:
  APPLICANT: Allen, Stephen
  TITLE OF INVENTION: Plant Cellulose Synthases
  FILE REFERENCE: BB1170 US CIP
  CURRENT APPLICATION NUMBER: US/09/900,237
  CURRENT FILING DATE:  2001-07-06
  PRIOR APPLICATION NUMBER: 60/092,844
  PRIOR FILING DATE: 1998-07-14
  PRIOR APPLICATION NUMBER: PCT/US99/15871
  PRIOR FILING DATE: 1999-07-13
  PRIOR APPLICATION NUMBER: 09/720383
  PRIOR FILING DATE: 2000-12-21
  NUMBER OF SEQ ID NOS: 33
  SOFTWARE: Microsoft Office 97
 SEQ ID NO 7
   LENGTH: 3786
   TYPE: DNA
   ORGANISM: Zea mays
S-09-900-237-7

 Query Match             41.1%;  Score 1416.4;  DB 10;  Length 3786;
 Best Local Similarity   66.7%;  Pred. No. 0;
 Matches 2128;  Conservative    0;  Mismatches 1026;  Indels   36;  Gaps    6;

y       103 GACATGTCAAATCTGCAGAGATGAGATCGAATTGACTGTTGATGGAGAACCGTTTGTGGC 162
o       331 GCCGTGCCAGATATGCGGCGACGAGGTCGGGGTGGGCTTCGACGGGGAGCCCTTCGTGGC 390

Qy      163 ATGTAACGAATGTGCATTCCCTGTGTGTAGACCTTGCTATGAGTACGAAAGACGAGAAGG 222
Db      391 GTGCAACGAGTGCGCCTTCCCCGTCTGCCGCGCCTGCTACGAGTACGAGCGCCGCGAGGG 450

Qy      223 CAATCAAGCTTGTCCACAGTGCAAAACCCGTTTCAAACGTCTTAAAGGAAGTCCAAGAGT 282
Db      451 CTCGCAAGCGTGCCCGCAGTGCAGGACCCGCTACAAGCGCCTCAAGGGCTGCCCGCGGGT 510

Qy      283 TGAAGGTGATGAAGAGGAAGATGACATTGATGATTTAGACAATGAGTT--------TGA 333
Db      511 GGCCGGCGACGAGGAGGAGGACGGCGTCGACGACCTGGAGGGCGAGTTCGGCCTGCAGGA 570

Qy      334 GTATGGAAATAATGGGATTGGATTTGATCAGGTTTCTGAAGGTATGTCAATCTCTCGTCG 393
Db      571 CGGCGCCGCCCACGAGGACGACCCGCAGTACGTCGCCGAGTCCATGCTCAGGGCGCAGAT 630

Qy      394 CAACTCCGGTTTCCCACAATCTGATTTGGATTCAGCTCCACCTGGCTCTCAGATTCCATT 453
Db      631 GAGCTACGGC[illegible]TTCAGCCCCGTCCCCAACGTGCCGCT 690

Qy      454 GCTGACTTAC[illegible]AGGAGGACGTTGAGATTT[illegible]TCTGATAGACATGCTCTTATTGTTCC 513
Db      691 CCTCACCAAC[illegible]AGATGGTTGATGACATC[illegible]GGAGCAGCACGCGCTCGTGCCGTC 750

Qy      514 TCCTTCACTTGGTGGTCAT---GGCAATAGAGTTCATCCTGTTTCTCTTTCTGACCC 567
Db      751 CTACATGAGCGGCGGCGGCGGCGGGGGCAAGAGGATCCACCCGCTCCCTTTCGCAGATCC 810

Qy      568 GACCGTGGCTGCACATCGAAGGCTGATGGTACCTCAGAAAGATCTTGCGGTTTATGGTTA 627
Db      811 CAACCTTCCAGTGCAACCGAGATCCATGGACCGGTCCAAGGATCTGGCCGCCTACGGATA 870

Qy      628 TGGAAGTGTCGCTTGGAAAGATCCGATGGAGGAATGGAAGAGAAAGCAGAATGAGAAACT 687
Db      871 TGGCAGCGTGGCCTGGAAGGAGAGAATGGAGGCTGGAAGCAGAAGCAG---GAGCGCCT 927

Qy      688 TCAGGTTGTTAGGCATGAAGGAGATCCTGATTTTGAAGATGGTGATGATGCTGATTTTCC 747
Db      928 GCAGCATGTCAGGAGCGAGGGTGGCGGTGATTGG---GATGGCGACGATGCAGATCTGCC 984

Qy      748 AATGATGGATGAGGGAAGGCAGCCATTGTCTATGAAGATACCAATCAAATCGAGCAAGAT 807
Db      985 ACTAATGGATGAAGCTAGGCAGCCATTGTCCAGAAAAGTCCCTATATCATCAAGCCGAAT 1044

Qy      808 AAATCCTTACCGGATGTTAATTGTGCTACGTCTTGTGATTCTTGGTCTCTTCTTTCACTA 867
Db      1045 TAATCCCTACAGGATGATTATCGTTATCCGGTTGGTGGTTTTGGGTTTCTTCTCCACTA 1104

Qy      868 CCGTATTCTTCACCCCGTCAAAGATGCATATGCTTTGTGGCTTATTTCTGTTATATGTGA 927
Db      1105 CCGAGTGATGCATCCGGCGAAAGATGCATTTGCATTGTGGCTCATATCTGTAATCTGTGA 1164

Qy      928 GATATGGTTTGCTGTTTCATGGGTTCTTGATCAGTTCCCTAAATGGTACCCTATCGAGCG 987
Db      1165 AATCTGGTTTGCGATGTCCTGGATTCTTGATCAGTTCCCAAAGTGGCTTCCAATCGAGAG 1224

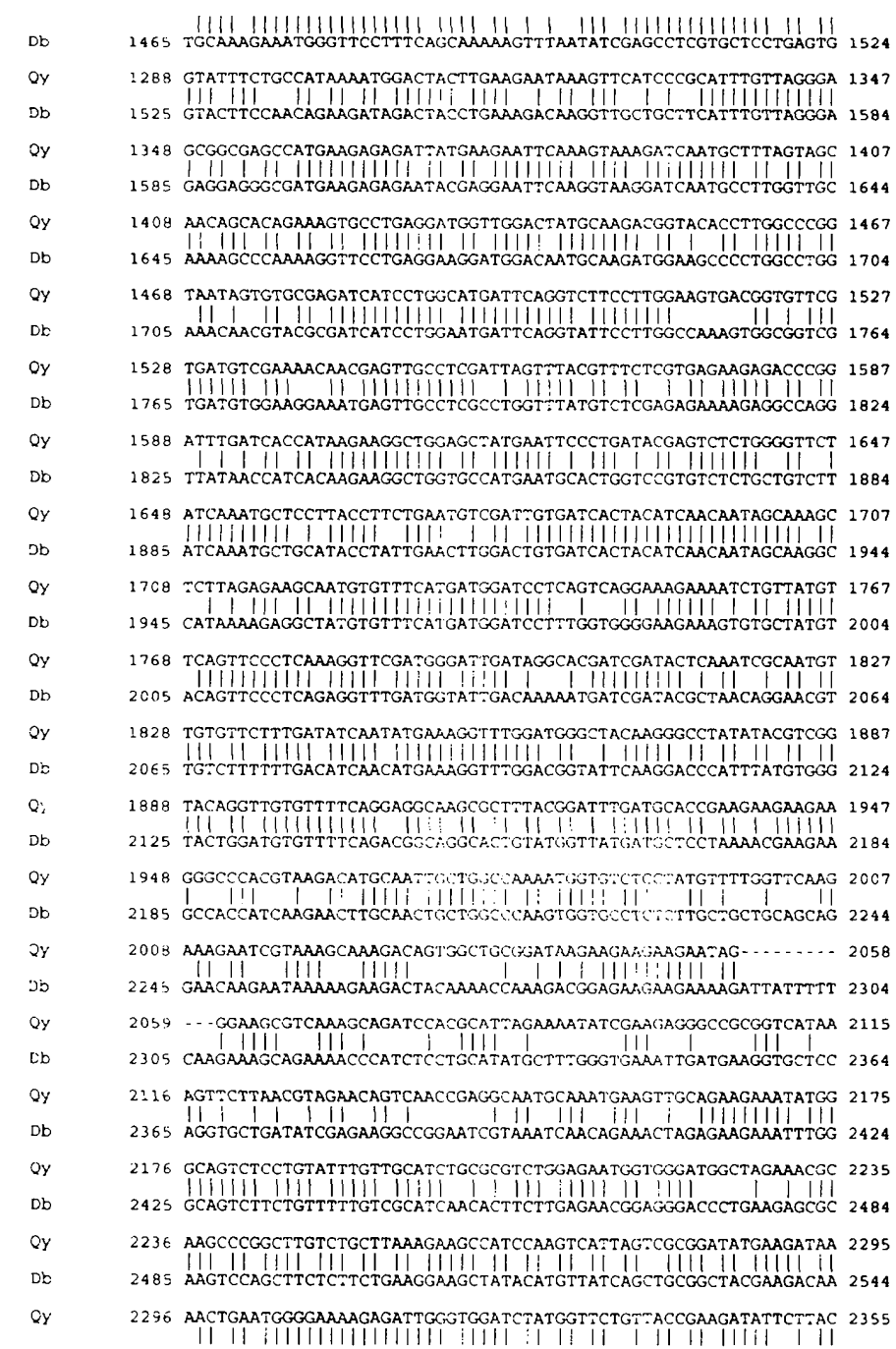
Qy      988 AGAAACGTACTTGGACCGACTCTCATTAAGATATGAGAAAGAAGGGAAACCGTCGGGACT 1047
Db      1225 AGAGACTTACCTGGACCGTTTGTCACTAAGGTTTGACAAGGAAGGTCAACCCTCTCAGCT 1284

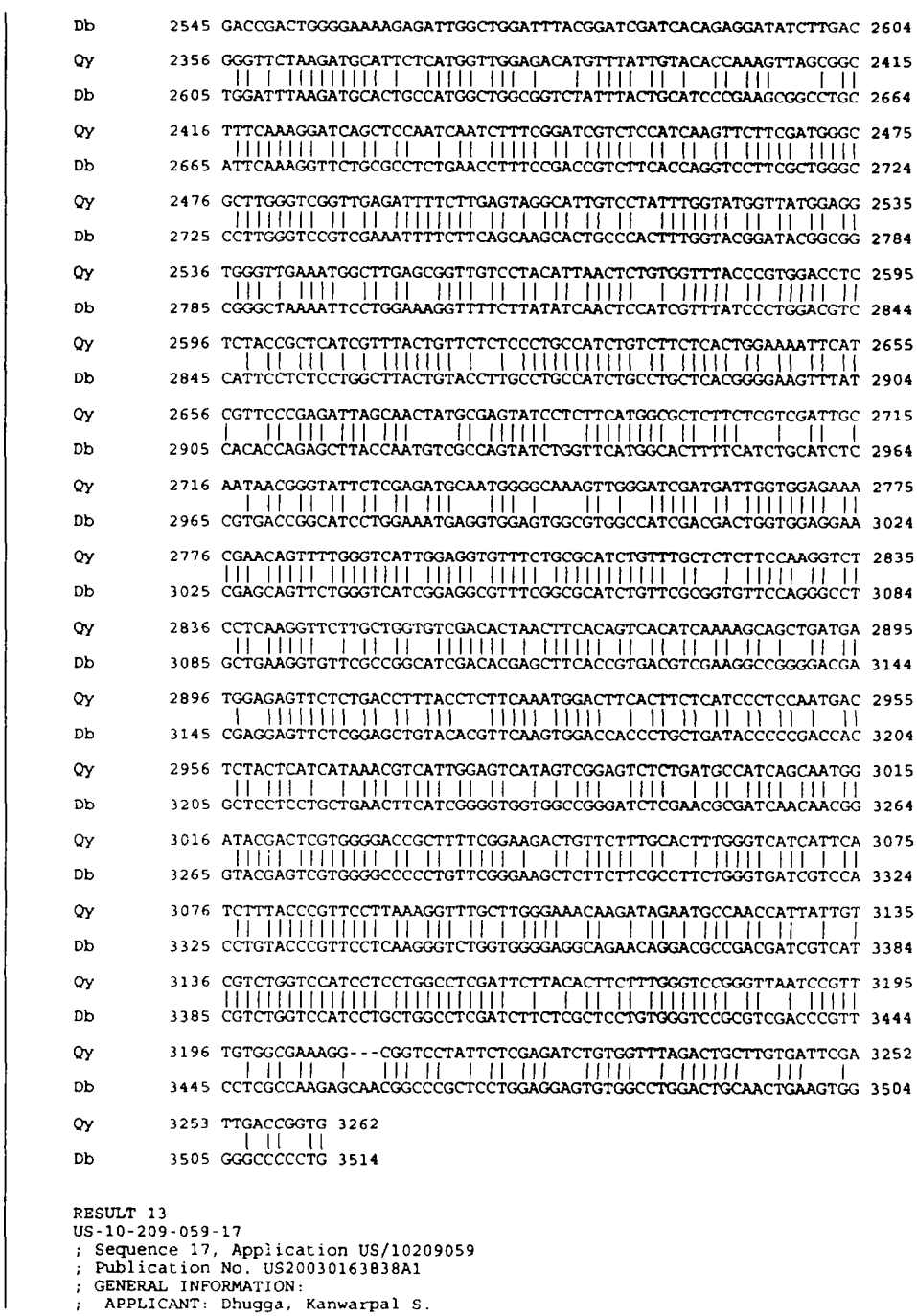
Qy      1048 ATCCCCTGTGGATGTATTTGTTAGTACAGTGGATCCATTGAAAGAGCCTCCGCTTATTAC 1107
Db      1285 TGCTCCAATCGACTTCTTTGTCAGTACGGTTGATCCCACAAAGGAACCTCCCTTGGTCAC 1344

Qy      1108 TGCAAATACTGTCTTGTCTATTCTTGCTGTTGATTATCCTGTCGATAAGGTTGCTTGTTA 1167
Db      1345 AGCGAACACTGTCCTTTCCATCCTTTCTGTGGATTATCCGGTTGAGAAGGTCTCCTGCTA 1404

Qy      1168 CGTATCTGATGATGGTGCTGCTATGCTTACTTTCGAAGCTCTTTCTGAGACCGCTGAATT 1227
Db      1405 TGTTTCTGATGATGGTGCTGCAATGCTTACGTTTGAAGCATTGTCTGAAACATCTGAATT 1464

Qy      1228 CGCAAGGAAATGGGTTCCTTTCTGCAAGAAATATTGTATTGAGCCTCGTGCTCCCGAATG 1287
```





RESULT 13
US-10-209-059-17
; Sequence 17, Application US/10209059
; Publication No. US20030163838A1
; GENERAL INFORMATION:
; APPLICANT: Dhugga, Kanwarpal S.

APPLICANT: Wang, Haiyin
TITLE OF INVENTION: Maize Cellulose Synthases and Uses
TITLE OF INVENTION: Thereof
FILE REFERENCE: 0864R2
CURRENT APPLICATION NUMBER: US/10/209,059
CURRENT FILING DATE: 2002-07-31
PRIOR APPLICATION NUMBER: 60/096,822
PRIOR FILING DATE: 1998-08-17
PRIOR APPLICATION NUMBER: 09/371,383
PRIOR FILING DATE: 1999-08-06
PRIOR APPLICATION NUMBER: 09/550,483
PRIOR FILING DATE: 2000-04-14
NUMBER OF SEQ ID NOS: 52
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 17
LENGTH: 3813
TYPE: DNA
ORGANISM: Zea mays
JS-10-209-059-17

Query Match 41.1%; Score 1416.4; DB 12; Length 3813;
Best Local Similarity 66.7%; Pred. No. 0;
Matches 2128; Conservative 0; Mismatches 1026; Indels 36; Gaps 6;

```
Qy      103 GACATGTCAAATCTGCAGAGATGAGATCGAATTGACTGTTGATGGAGAACCGTTTGTGGC 162
Db      331 GCCGTGCCAGATATGCGGCGACGAGGTCGGGGTGGGCTTCGACGGGGAGCCCTTCGTGGC 390

Qy      163 ATGTAACGAATGTGCATTCCCTGTGTGTAGACCTTGCTATGAGTACGAAAGACGAGAAGG 222
Db      391 GTGCAACGAGTGCGCCTTCCCCGTCTGCCGCGCCTGCTACGAGTACGAGCGCCGCGAGGG 450

Qy      223 CAATCAAGCTTGTCCACAGTGCAAAACCCGTTTCAAACGTCTTAAAGGAAGTCCAAGAGT 282
Db      451 CTCGCAAGCGTGCCCGCAGTGCAGGACCCGCTACAAGCGCCTCAAGGGCTGCCCGCGGGT 510

Qy      283 TGAAGGTGATGAAGAGGAAGATGACATTGATGATTTAGACAATGAGTT---------TGA 333
Db      511 GGCCGGCGACGAGGAGGAGGACGGCGTCGACGACCTGGAGGGCGAGTTCGGCCTGCAGGA 570

Qy      334 GTATGGAAATAATGGGATTGGATTTGATCAGGTTTCTGAAGGTATGTCAATCTCTCGTCG 393
Db      571 CGGCGCCGCCCACGAGGACGACCCGCAGTACGTCGCCGAGTCCATGCTCAGGGCGCAGAT 630

Qy      394 CAACTCCGGTTTCCCACAATCTGATTTGGATTCAGCTCCACCTGGCTCTCAGATTCCATT 453
Db      631 GAGCTACGGCCGCGGCGGCGACGCGCACCCCGGCTTCAGCCCCGTCCCCAACGTGCCGCT 690

Qy      454 GCTGACTTACGGCGACGAGGACGTTGAGATTTCTTCTGATAGACATGCTCTTATTGTTCC 513
Db      691 CCTCACCAACGGCCAGATGGTTGATGACATCCCGCCGGAGCAGCACGCGCTCGTGCCGTC 750

Qy      514 TCCTTCACTTGGTGGTCAT------GGCAATAGAGTTCATCCTGTTTCTCTTTCTGACCC 567
Db      751 CTACATGAGCGGCGGCGGCGGCGGGGGCAAGAGGATCCACCCGCTCCCTTTCGCAGATCC 810

Qy      568 GACCGTGGCTGCACATCGAAGGCTGATGGTACCTCAGAAAGATCTTGCGGTTTATGGTTA 627
Db      811 CAACCTTCCAGTGCAACCGAGATCCATGGACCCGTCCAAGGATCTGGCCGCCTACGGATA 870

Qy      628 TGGAAGTGTCGCTTGGAAAGATCGGATGGAGGAATGGAAGAGAAAGCAGAATGAGAAACT 687
Db      871 TGGCAGCGTGGCCTGGAAGGAGAGAATGGAGGGCTGGAAGCAGAAGCAG---GAGCGCCT 927

Qy      688 TCAGGTTGTTAGGCATGAAGGAGATCCTGATTTTGAAGATGGTGATGATGCTGATTTTCC 747
Db      928 GCAGCATGTCAGGAGCGAGGGTGGCGGTGATTGG---GATGGCGACGATGCAGATCTGCC 984

Qy      748 AATGATGGATGAGGGAAGGCAGCCATTGTCTATGAAGATACCAATCAAATCGAGCAAGAT 807
Db      985 ACTAATGGATGAAGCTAGGCAGCCATTGTCCAGAAAAGTCCCTATATCATCAAGCCGAAT 1044

Qy      808 AAATCCTTACCGGATGTTAATTGTGCTACGTCTTGTGATTCTTGGTCTCTTCTTTCACTA 867
```

```
Db      1045 TAATCCCTACAGGATGATTATCGTTATCCGGTTGGTGGTTTTGGGTTTCTTCTTCCACTA 1104

Qy       868 CCGTATTCTTCACCCCGTCAAAGATGCATATGCTTTGTGGCTTATTTCTGTTATATGTGA 927
Db      1105 CCGAGTGATGCATCCGGCGAAAGATGCATTTGCATTGTGGCTCATATCTGTAATCTGTGA 1164

Qy       928 GATATGGTTTGCTGTTTCATGGGTTCTTGATCAGTTCCCTAAATGGTACCCTATCGAGCG 987
Db      1165 AATCTGGTTTGCGATGTCCTGGATTCTTGATCAGTTCCCAAAGTGGCTTCCAATCGAGAG 1224

Qy       988 AGAAACGTACTTGGACCGACTCTCATTAAGATATGAGAAAGAAGGGAAACCGTCGGGACT 1047
Db      1225 AGAGACTTACCTGGACCGTTTGTCACTAAGGTTTGACAAGGAAGGTCAACCCTCTCAGCT 1284

Qy      1048 ATCCCCTGTGGATGTATTTGTTAGTACAGTGGATCCATTGAAAGAGCCTCCGCTTATTAC 1107
Db      1285 TGCTCCAATCGACTTCTTTGTCAGTACGGTTGATCCCACAAAGGAACCTCCCTTGGTCAC 1344

Qy      1108 TGCAAATACTGTCTTGTCTATTCTTTCTGTTGATTATCCTGTCGATAAGGTTGCTTGTTA 1167
Db      1345 AGCGAACACCTCTCTTTCCATCCTTTCTGTGGATTATCCGGTTGAGAAGGTCTCCTGCTA 1404

Qy      1168 CGTATCTGATGATGGTGCTGCTATGCTTACTTTCGAAGCTCTTTCTGAGACCGCTGAATT 1227
Db      1405 TGTTTCTGATGATGGTGCTGCAATGCTTACGTTTGAAGCATTGTCTGAAACATCTGAATT 1464

Qy      1228 CGCAAGGAAATGGGTTCCTTTCTGCAAGAAATATTGTATTGAGCCTCGTGCTCCCGAATG 1287
Db      1465 TGCAAAGAAATGGGTTCCTTTCAGCAAAAAGTTTAATATCGAGCCTCGTGCTCCTGAGTG 1524

Qy      1288 GTATTTCTGCCATAAAATGGACTACTTGAAGAATAAAGTTCATCCCGCATTTGTTAGGGA 1347
Db      1525 GTACTTCCAACAGAAGATAGACTACCTGAAAGACAAGGTTGCTGCTTCATTTGTTAGGGA 1584

Qy      1348 GCGGCGAGCCATGAAGAGAGATTATGAAGAATTCAAAGTAAAGATCAATGCTTTAGTAGC 1407
Db      1585 GAGGAGGGCGATGAAGAGAGAATACGAGGAATTCAAGGTAAGGATCAATGCCTTGGTTGC 1644

Qy      1408 AACAGCACAGAAAGTGCCTGAGGATGGTTGGACTATGCAAGACGGTACACCTTGGCCCGG 1467
Db      1645 AAAAGCCCAAAAGGTTCCTGAGGAAGGATGGACAATGCAAGATGGAAGCCCCTGGCCTGG 1704

Qy      1468 TAATAGTGTGCGAGATCATCCTGGCATGATTCAGGTCTTCCTTGGAAGTGACGGTGTTCG 1527
Db      1705 AAACAACGTACGCGATCATCCTGGAATGATTCAGGTATTCCTTGGCCAAAGTGGCGGTCG 1764

Qy      1528 TGATGTCGAAAACAACGAGTTGCCTCGATTAGTTTACGTTTCTCGTGAGAAGAGACCCGG 1587
Db      1765 TGATGTGGAAGGAAATGAGTTGCCTCGCCTGGTTTATGTCTCGAGAGAAAAGAGGCCAGG 1824

Qy      1588 ATTTGATCACCATAAGAAGGCTGGAGCTATGAATTCCCTGATACGAGTCTCTGGGGTTCT 1647
Db      1825 TTATAACCATCACAAGAAGGCTGGTGCCATGAATGCACTGGTCCGTGTCTCTGCTGTCTT 1884

Qy      1648 ATCAAATGCTCCTTACCTTCTGAATGTCGATTGTGATCACTACATCAACAATAGCAAAGC 1707
Db      1885 ATCAAATGCTGCATACCTATTGAACTTGGACTGTGATCACTACATCAACAATAGCAAGGC 1944

Qy      1708 TCTTAGAGAAGCAATGTGTTTCATGATGGATCCTCAGTCAGGAAAGAAAATCTGTTATGT 1767
Db      1945 CATAAAAGAGGCTATGTGTTTCATGATGGATCCTTTGGTGGGGAAGAAAGTGTGCTATGT 2004

Qy      1768 TCAGTTCCCTCAAAGGTTCGATGGGATTGATAGGCACGATCGATACTCAAATCGCAATGT 1827
Db      2005 ACAGTTCCCTCAGAGGTTTGATGGTATTGACAAAAATGATCGATACGCTAACAGGAACGT 2064

Qy      1828 TGTGTTCTTTGATATCAATATGAAAGGTTTGGATGGGCTACAAGGGCCTATATACGTCGG 1887
Db      2065 TGTCTTTTTTGACATCAACATGAAAGGTTTGGACGGTATTCAAGGACCCATTTATGTGGG 2124

Qy      1888 TACAGGTTGTGTTTTCAGGAGGCAAGCGCTTTACGGATTTGATGCACCGAAGAAGAAGAA 1947
```

```
Db       2125 TACTGGATGTGTTTTCAGACGGCAGGCACTGTATGGTTATGATGCTCCTAAAACGAAGAA 2184

Qy       1948 GGGCCCACGTAAGACATGCAATTGCTGGCCAAAATGGTGTCTCCTATGTTTTGGTTCAAG 2007

Db       2185 GCCACCATCAAGAACTTGCAACTGCTGGCCCAAGTGGTGCCTCTCTTGCTGCTGCAGCAG 2244

Qy       2008 AAAGAATCGTAAAGCAAAGACAGTGGCTGCGGATAAGAAGAAGAAGAATAG--------- 2058

Db       2245 GAACAAGAATAAAAAGAAGACTACAAAACCAAAGACGGAGAAGAAGAAAAGATTATTTTT 2304

Qy       2059 ---GGAAGCGTCAAAGCAGATCCACGCATTAGAAAATATCGAAGAGGGCCGCGGTCATAA 2115

Db       2305 CAAGAAAGCAGAAAACCCATCTCCTGCATATGCTTTGGGTGAAATTGATGAAGGTGCTCC 2364

Qy       2116 AGTTCTTAACGTAGAACAGTCAACCGAGGCAATGCAAATGAAGTTGCAGAAGAAATATGG 2175

Db       2365 AGGTGCTGATATCGAGAAGGCCGGAATCGTAAATCAACAGAAACTAGAGAAGAAATTTGG 2424

Qy       2176 GCAGTCTCCTGTATTTGTTGCATCTGCGCGTCTGGAGAATGGTGGGATGGCTAGAAACGC 2235

Db       2425 GCAGTCTTCTGTTTTTGTCGCATCAACACTTCTTGAGAACGGAGGGACCCTGAAGAGCGC 2484

Qy       2236 AAGCCCGGCTTGTCTGCTTAAAGAAGCCATCCAAGTCATTAGTCGCGGATATGAAGATAA 2295

Db       2485 AAGTCCAGCTTCTCTTCTGAAGGAAGCTATACATGTTATCAGCTGCGGCTACGAAGACAA 2544

Qy       2296 AACTGAATGGGGAAAAGAGATTGGGTGGATCTATGGTTCTGTTACCGAAGATATTCTTAC 2355

Db       2545 GACCGACTGGGGAAAAGAGATTGGCTGGATTTACGGATCGATCACAGAGGATATCTTGAC 2604

Qy       2356 GGGTTCTAAGATGCATTCTCATGGTTGGAGACATGTTTATTGTACACCAAAGTTAGCGGC 2415

Db       2605 TGGATTTAAGATGCACTGCCATGGCTGGCGGTCTATTTACTGCATCCCGAAGCGGCCTGC 2664

Qy       2416 TTTCAAAGGATCAGCTCCAATCAATCTTTCGGATCGTCTCCATCAAGTTCTTCGATGGGC 2475

Db       2665 ATTCAAAGGTTCTGCGCCTCTGAACCTTTCCGACCGTCTTCACCAGGTCCTTCGCTGGGC 2724

Qy       2476 GCTTGGGTCGGTTGAGATTTTCTTGAGTAGGCATTGTCCTATTTGGTATGGTTATGGAGG 2535

Db       2725 CCTTGGGTCCGTCGAAATTTTCTTCAGCAAGCACTGCCCACTTTGGTACGGATACGGCGG 2784

Qy       2536 TGGGTTGAAATGGCTTGAGCGGTTGTCCTACATTAACTCTGTGGTTTACCCGTGGACCTC 2595

Db       2785 CGGGCTAAAATTCCTGGAAAGGTTTTCTTATATCAACTCCATCGTTTATCCCTGGACGTC 2844

Qy       2596 TCTACCGCTCATCGTTTACTGTTCTCTCCCTGCCATCTGTCTTCTCACTGGAAAATTCAT 2655

Db       2845 CATTCCTCTCCTGGCTTACTGTACCTTGCCTGCCATCTGCCTGCTCACGGGGAAGTTTAT 2904

Qy       2656 CGTTCCCGAGATTAGCAACTATGCGAGTATCCTCTTCATGGCGCTCTTCTCGTCGATTGC 2715

Db       2905 CACACCAGAGCTTACCAATGTCGCCAGTATCTGGTTCATGGCACTTTTCATCTGCATCTC 2964

Qy       2716 AATAACGGGTATTCTCGAGATGCAATGGGGCAAAGTTGGGATCGATGATTGGTGGAGAAA 2775

Db       2965 CGTGACCGGCATCCTGGAAATGAGGTGGAGTGGCGTGGCCATCGACGACTGGTGGAGGAA 3024

Qy       2776 CGAACAGTTTTGGGTCATTGGAGGTGTTTCTGCGCATCTGTTTGCTCTCTTCCAAGGTCT 2835

Db       3025 CGAGCAGTTCTGGGTCATCGGAGGCGTTTCGGCGCATCTGTTCGCGGTGTTCCAGGGCCT 3084

Qy       2836 CCTCAAGGTTCTTGCTGGTGTCGACACTAACTTCACAGTCACATCAAAAGCAGCTGATGA 2895

Db       3085 GCTGAAGGTGTTCGCCGGCATCGACACGAGCTTCACCGTGACGTCGAAGGCCGGGGACGA 3144

Qy       2896 TGGAGAGTTCTCTGACCTTTACCTCTTCAAATGGACTTCACTTCTCATCCCTCCAATGAC 2955

Db       3145 CGAGGAGTTCTCGGAGCTGTACACGTTCAAGTGGACCACCCTGCTGATACCCCCGACCAC 3204

Qy       2956 TCTACTCATCATAAACGTCATTGGAGTCATAGTCGGAGTCTCTGATGCCATCAGCAATGG 3015

Db       3205 GCTCCTCCTGCTGAACTTCATCGGGGTGGTGGCCGGGATCTCGAACGCGATCAACAACGG 3264

Qy       3016 ATACGACTCGTGGGGACCGCTTTTCGGAAGACTGTTCTTTGCACTTTGGGTCATCATTCA 3075

Db       3265 GTACGAGTCGTGGGGCCCCCTGTTCGGGAAGCTCTTCTTCGCCTTCTGGGTGATCGTCCA 3324

Qy       3076 TCTTTACCCGTTCCTTAAAGGTTTGCTTGGGAAACAAGATAGAATGCCAACCATTATTGT 3135

Db       3325 CCTGTACCCGTTCCTCAAGGGTCTGGTGGGGAGGCAGAACAGGACGCCGACGATCGTCAT 3384

Qy       3136 CGTCTGGTCCATCCTCCTGGCCTCGATTCTTACACTTCTTTGGGTCCGGGTTAATCCGTT 3195

Db       3385 CGTCTGGTCCATCCTGCTGGCCTCGATCTTCTCGCTCCTGTGGGTCCGCGTCGACCCGTT 3444

Qy       3196 TGTGGCGAAAGG---CGGTCCTATTCTCGAGATCTGTGGTTTAGACTGCTTGTGATTCGA 3252

Db       3445 CCTCGCCAAGAGCAACGGCCCGCTCCTGGAGGAGTGTGGCCTGGACTGCAACTGAAGTGG 3504

Qy       3253 TTGACCGGTG 3262

Db       3505 GGGCCCCCTG 3514


RESULT 14
US-10-160-719-25
; Sequence 25, Application US/10160719
; Publication No. US20030167528A1
; GENERAL INFORMATION:
;  APPLICANT: Kanwarpal S. Dhugga
;  APPLICANT:  Timothy G. Helentjaris
;  APPLICANT:  Benjamin A.  Bowen
;  APPLICANT:  Xun Wang
;  TITLE OF INVENTION: Maize Cellulose Synthases and Uses
;  TITLE OF INVENTION:  Thereof
;  FILE REFERENCE: 0864
;  CURRENT APPLICATION NUMBER: US/10/160,719
;  CURRENT FILING DATE:  2002-06-03
;  PRIOR APPLICATION NUMBER: US/09/371,383A
;  PRIOR FILING DATE: 1999-08-06
;  PRIOR APPLICATION NUMBER: US 60/096,822
;  PRIOR FILING DATE: 1998-08-17
;  NUMBER OF SEQ ID NOS: 60
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 25
;   LENGTH: 3813
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (215)...(3496)
US-10-160-719-25

  Query Match             41.1%;  Score 1416.4;  DB 12;  Length 3813;
  Best Local Similarity   66.7%;  Pred. No. 0;
  Matches 2128;  Conservative    0;  Mismatches 1026;  Indels   36;  Gaps    6;

Qy        103 GACATGTCAAATCTGCAGAGATGAGATCGAATTGACTGTTGATGGAGAACCGTTTGTGGC 162

Db        331 GCCGTGCCAGATATGCGGCGACGAGGTCGGGGTGGGCTTCGACGGGGAGCCCTTCGTGGC 390

Qy        163 ATGTAACGAATGTGCATTCCCTGTGTGTAGACCTTGCTATGAGTACGAAAGACGAGAAGG 222

Db        391 GTGCAACGAGTGCGCCTTCCCCGTCTGCCGCGCCTGCTACGAGTACGAGCGCCGCGAGGG 450

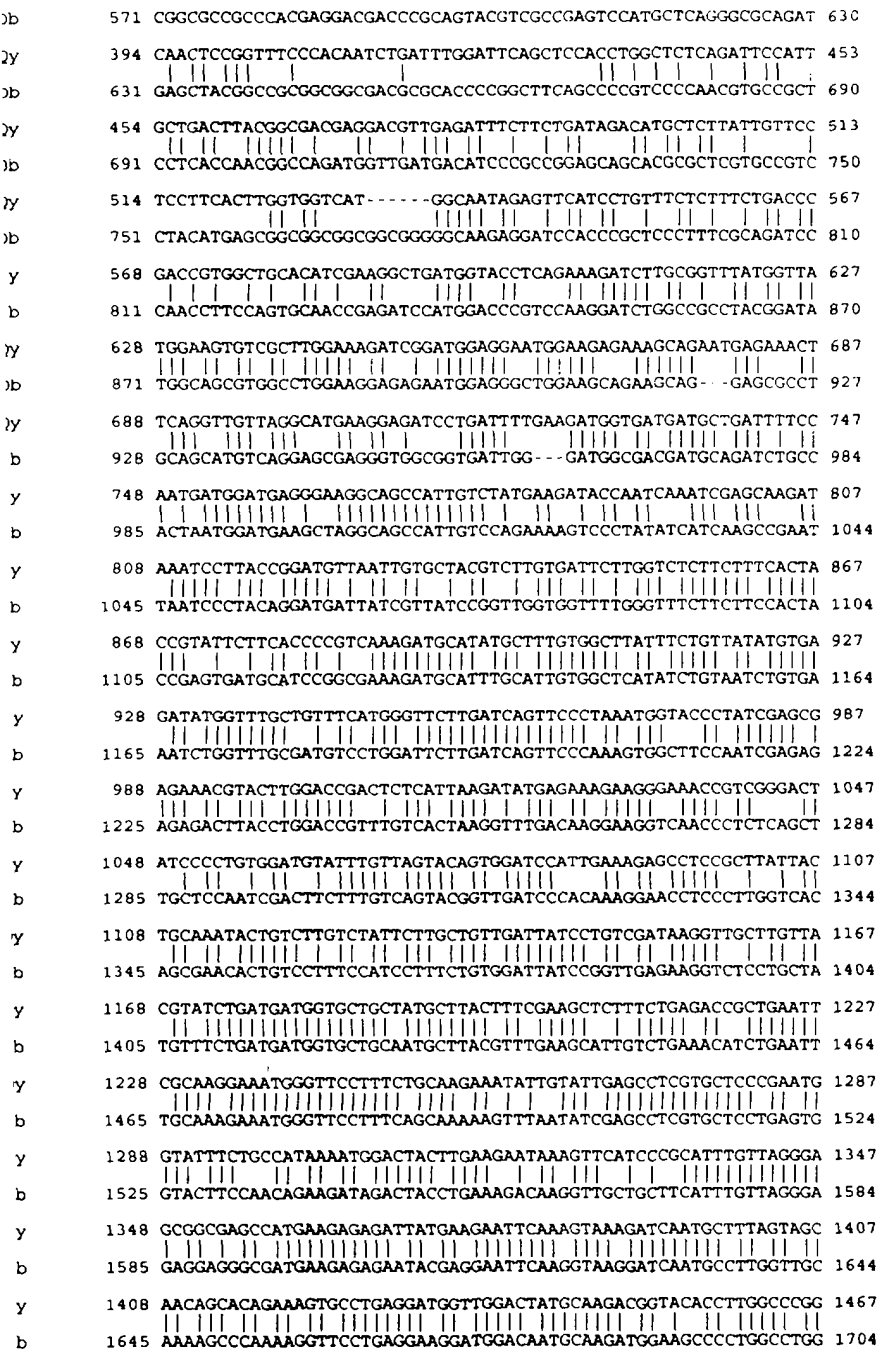
Qy        223 CAATCAAGCTTGTCCACAGTGCAAAACCCGTTTCAAACGTCTTAAAGGAAGTCCAAGAGT 282

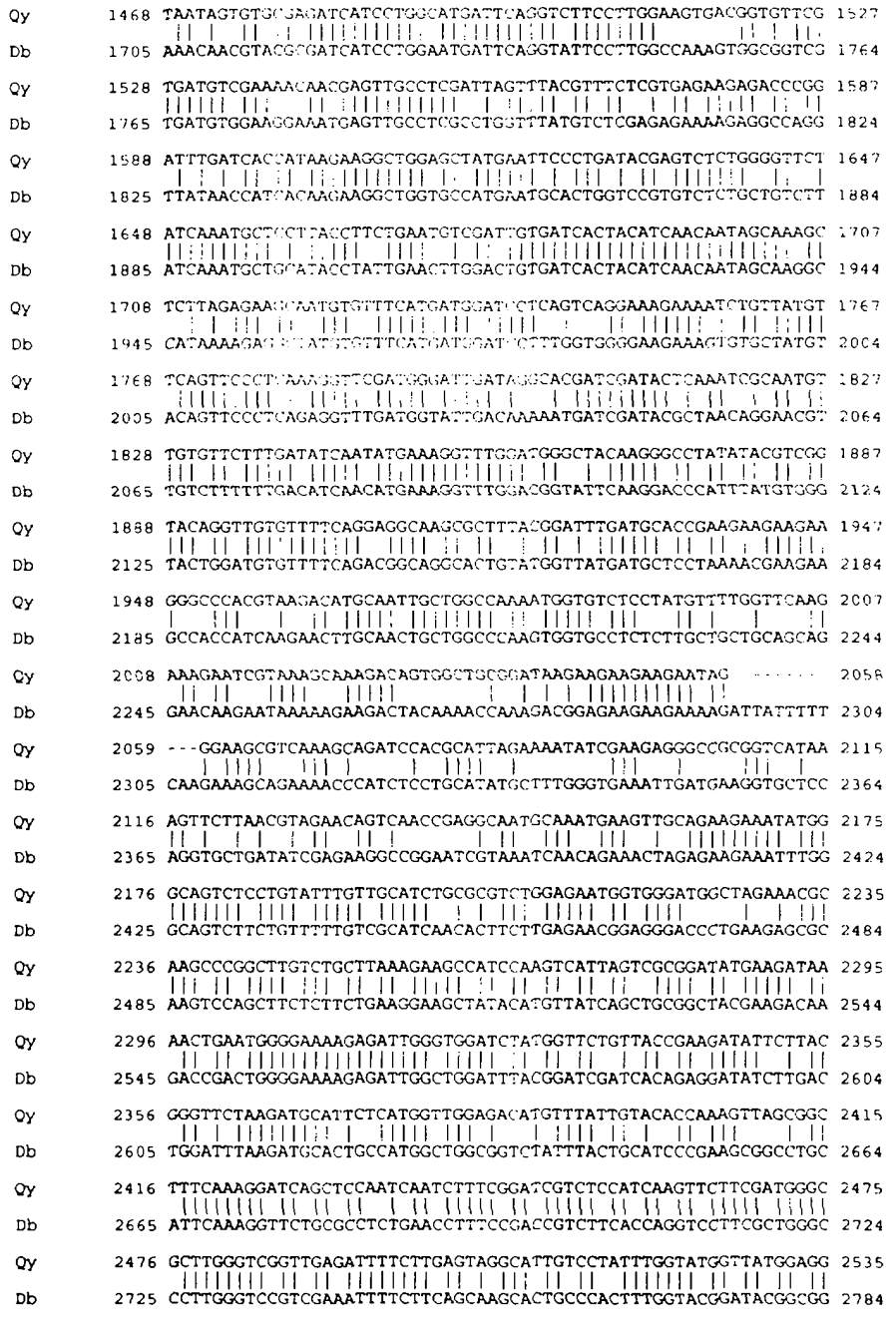
Db        451 CTCGCAAGCGTGCCCGCAGTGCAGGACCCGCTACAAGCGCCTCAAGGGCTGCCCGCGGGT 510

Qy        283 TGAAGGTGATGAAGAGGAAGATGACATTGATGATTTAGACAATGAGTT---------TGA 333

Db        511 GGCCGGCGACGAGGAGGAGGACGGCGTCGACGACCTGGAGGGCGAGTTCGGCCTGCAGGA 570

Qy        334 GTATGGAAATAATGGGATTGGATTTGATCAGGTTTCTGAAGGTATGTCAATCTCTCGTCG 393
```


Db 571 CGGCGCCGCCCACGAGGACGACCCGCAGTACGTCGCCGAGTCCATGCTCAGGGCGCAGAT 630
Qy 394 CAACTCCGGTTTCCCACAATCTGATTTGGATTCAGCTCCACCTGGCTCTCAGATTCCATT 453
Db 631 GAGCTACGGCCGCGGCGGCGACGCGCACCCCGGCTTCAGCCCCGTCCCCAACGTGCCGCT 690
Qy 454 GCTGACTTACGGCGACGAGGACGTTGAGATTTCTTCTGATAGACATGCTCTTATTGTTCC 513
Db 691 CCTCACCAACGGCCAGATGGTTGATGACATCCCGCCGGAGCAGCACGCGCTCGTGCCGTC 750
Qy 514 TCCTTCACTTGGTGGTCAT------GGCAATAGAGTTCATCCTGTTTCTCTTTCTGACCC 567
Db 751 CTACATGAGCGGCGGCGGCGGCGGGGGCAAGAGGATCCACCCGCTCCCTTTCGCAGATCC 810
Qy 568 GACCGTGGCTGCACATCGAAGGCTGATGGTACCTCAGAAAGATCTTGCGGTTTATGGTTA 627
Db 811 CAACCTTCCAGTGCAACCGAGATCCATGGACCCGTCCAAGGATCTGGCCGCCTACGGATA 870
Qy 628 TGGAAGTGTCGCTTGGAAAGATCGGATGGAGGAATGGAAGAGAAAGCAGAATGAGAAACT 687
Db 871 TGGCAGCGTGGCCTGGAAGGAGAGAATGGAGGGCTGGAAGCAGAAGCAG---GAGCGCCT 927
Qy 688 TCAGGTTGTTAGGCATGAAGGAGATCCTGATTTTGAAGATGGTGATGATGCTGATTTTCC 747
Db 928 GCAGCATGTCAGGAGCGAGGGTGGCGGTGATTGG---GATGGCGACGATGCAGATCTGCC 984
Qy 748 AATGATGGATGAGGGAAGGCAGCCATTGTCTATGAAGATACCAATCAAATCGAGCAAGAT 807
Db 985 ACTAATGGATGAAGCTAGGCAGCCATTGTCCAGAAAAGTCCCTATATCATCAAGCCGAAT 1044
Qy 808 AAATCCTTACCGGATGTTAATTGTGCTACGTCTTGTGATTCTTGGTCTCTTCTTTCACTA 867
Db 1045 TAATCCCTACAGGATGATTATCGTTATCCGGTTGGTGGTTTTGGGTTTCTTCTTCCACTA 1104
Qy 868 CCGTATTCTTCACCCCGTCAAAGATGCATATGCTTTGTGGCTTATTTCTGTTATATGTGA 927
Db 1105 CCGAGTGATGCATCCGGCGAAAGATGCATTTGCATTGTGGCTCATATCTGTAATCTGTGA 1164
Qy 928 GATATGGTTTGCTGTTTCATGGGTTCTTGATCAGTTCCCTAAATGGTACCCTATCGAGCG 987
Db 1165 AATCTGGTTTGCGATGTCCTGGATTCTTGATCAGTTCCCAAAGTGGCTTCCAATCGAGAG 1224
Qy 988 AGAAACGTACTTGGACCGACTCTCATTAAGATATGAGAAAGAAGGGAAACCGTCGGGACT 1047
Db 1225 AGAGACTTACCTGGACCGTTTGTCACTAAGGTTTGACAAGGAAGGTCAACCCTCTCAGCT 1284
Qy 1048 ATCCCCTGTGGATGTATTTGTTAGTACAGTGGATCCATTGAAAGAGCCTCCGCTTATTAC 1107
Db 1285 TGCTCCAATCGACTTCTTTGTCAGTACGGTTGATCCCACAAAGGAACCTCCCTTGGTCAC 1344
Qy 1108 TGCAAATACTGTCTTGTCTATTCTTGCTGTTGATTATCCTGTCGATAAGGTTGCTTGTTA 1167
Db 1345 AGCGAACACTGTCCTTTCCATCCTTTCTGTGGATTATCCGGTTGAGAAGGTCTCCTGCTA 1404
Qy 1168 CGTATCTGATGATGGTGCTGCTATGCTTACTTTCGAAGCTCTTTCTGAGACCGCTGAATT 1227
Db 1405 TGTTTCTGATGATGGTGCTGCAATGCTTACGTTTGAAGCATTGTCTGAAACATCTGAATT 1464
Qy 1228 CGCAAGGAAATGGGTTCCTTTCTGCAAGAAATATTGTATTGAGCCTCGTGCTCCCGAATG 1287
Db 1465 TGCAAAGAAATGGGTTCCTTTCAGCAAAAAGTTTAATATCGAGCCTCGTGCTCCTGAGTG 1524
Qy 1288 GTATTTCTGCCATAAAATGGACTACTTGAAGAATAAAGTTCATCCCGCATTTGTTAGGGA 1347
Db 1525 GTACTTCCAACAGAAGATAGACTACCTGAAAGACAAGGTTGCTGCTTCATTTGTTAGGGA 1584
Qy 1348 GCGGCGAGCCATGAAGAGAGATTATGAAGAATTCAAAGTAAAGATCAATGCTTTAGTAGC 1407
Db 1585 GAGGAGGGCGATGAAGAGAGAATACGAGGAATTCAAGGTAAGGATCAATGCCTTGGTTGC 1644
Qy 1408 AACAGCACAGAAAGTGCCTGAGGATGGTTGGACTATGCAAGACGGTACACCTTGGCCCGG 1467
Db 1645 AAAAGCCCAAAAGGTTCCTGAGGAAGGATGGACAATGCAAGATGGAAGCCCCTGGCCTGG 1704


Qy 1468 TAATAGTGTGCGAGATCATCCTGGCATGATTCAGGTCTTCCTTGGAAGTGACGGTGTTCG 1527
Db 1705 AAACAACGTACGCGATCATCCTGGAATGATTCAGGTATTCCTTGGCCAAAGTGGCGGTCG 1764
Qy 1528 TGATGTCGAAAACAACGAGTTGCCTCGATTAGTTTACGTTTCTCGTGAGAAGAGACCCGG 1587
Db 1765 TGATGTGGAAGGAAATGAGTTGCCTCGCCTGGTTTATGTCTCGAGAGAAAAGAGGCCAGG 1824
Qy 1588 ATTTGATCACCATAAGAAGGCTGGAGCTATGAATTCCCTGATACGAGTCTCTGGGGTTCT 1647
Db 1825 TTATAACCATCACAAGAAGGCTGGTGCCATGAATGCACTGGTCCGTGTCTCTGCTGTCTT 1884
Qy 1648 ATCAAATGCTCCTTACCTTCTGAATGTCGATTGTGATCACTACATCAACAATAGCAAAGC 1707
Db 1885 ATCAAATGCTGCATACCTATTGAACTTGGACTGTGATCACTACATCAACAATAGCAAGGC 1944
Qy 1708 TCTTAGAGAAGCAAATGTGTTTCATGATGGATCTCAGTCAGGAAAGAAAATCTGTTATGT 1767
Db 1945 CATAAAAGAGCAATGTGTTTCATGATGGATCTTTGGTGGGGAAGAAAGTGTGCTATGT 2004
Qy 1768 TCAGTTCCCTCAAAGGTTCGATGGGATTGATAGGCACGATCGATACTCAAATCGCAATGT 1827
Db 2005 ACAGTTCCCTCAGAGGTTTGATGGTATTGACAAAAATGATCGATACGCTAACAGGAACGT 2064
Qy 1828 TGTGTTCTTTGATATCAATATGAAAGGTTTGGATGGGCTACAAGGGCCTATATACGTCGG 1887
Db 2065 TGTCTTTTTTGACATCAACATGAAAGGTTTGGACGGTATTCAAGGACCCATTTATGTGGG 2124
Qy 1888 TACAGGTTGTGTTTTCAGGAGGCAAGCGCTTTACGGATTTGATGCACCGAAGAAGAAGAA 1947
Db 2125 TACTGGATGTGTTTTCAGACGGCAGGCACTGTATGGTTATGATGCTCCTAAAACGAAGAA 2184
Qy 1948 GGGCCCACGTAAGACATGCAATTGCTGGCCAAAATGGTGTCTCCTATGTTTTGGTTCAAG 2007
Db 2185 GCCACCATCAAGAACTTGCAACTGCTGGCCCAAGTGGTGCCTCTCTTGCTGCTGCAGCAG 2244
Qy 2008 AAAGAATCGTAAAGCAAAGACAGTGGCTGCGGATAAGAAGAAGAAGAATAG ------ 2058
Db 2245 GAACAAGAATAAAAAGAAGACTACAAAACCAAAGACGGAGAAGAAGAAAAGATTATTTTT 2304
Qy 2059 ---GGAAGCGTCAAAGCAGATCCACGCATTAGAAAATATCGAAGAGGGCCGCGGTCATAA 2115
Db 2305 CAAGAAAGCAGAAAACCCATCTCCTGCATATGCTTTGGGTGAAATTGATGAAGGTGCTCC 2364
Qy 2116 AGTTCTTAACGTAGAACAGTCAACCGAGGCAATGCAAATGAAGTTGCAGAAGAAATATGG 2175
Db 2365 AGGTGCTGATATCGAGAAGGCCGGAATCGTAAATCAACAGAAACTAGAGAAGAAATTTGG 2424
Qy 2176 GCAGTCTCCTGTATTTGTTGCATCTGCGCGTCTGGAGAATGGTGGGATGGCTAGAAACGC 2235
Db 2425 GCAGTCTTCTGTTTTTGTCGCATCAACACTTCTTGAGAACGGAGGGACCCTGAAGAGCGC 2484
Qy 2236 AAGCCCGGCTTGTCTGCTTAAAGAAGCCATCCAAGTCATTAGTCGCGGATATGAAGATAA 2295
Db 2485 AAGTCCAGCTTCTCTTCTGAAGGAAGCTATACATGTTATCAGCTGCGGCTACGAAGACAA 2544
Qy 2296 AACTGAATGGGGAAAAGAGATTGGGTGGATCTATGGTTCTGTTACCGAAGATATTCTTAC 2355
Db 2545 GACCGACTGGGGAAAAGAGATTGGCTGGATTTACGGATCGATCACAGAGGATATCTTGAC 2604
Qy 2356 GGGTTCTAAGATGCATTCTCATGGTTGGAGACATGTTTATTGTACACCAAAGTTAGCGGC 2415
Db 2605 TGGATTTAAGATGCACTGCCATGGCTGGCGGTCTATTTACTGCATCCCGAAGCGGCCTGC 2664
Qy 2416 TTTCAAAGGATCAGCTCCAATCAATCTTTCGGATCGTCTCCATCAAGTTCTTCGATGGGC 2475
Db 2665 ATTCAAAGGTTCTGCGCCTCTGAACCTTTCCGACCGTCTTCACCAGGTCCTTCGCTGGGC 2724
Qy 2476 GCTTGGGTCGGTTGAGATTTTCTTGAGTAGGCATTGTCCTATTTGGTATGGTTATGGAGG 2535
Db 2725 CCTTGGGTCCGTCGAAATTTTCTTCAGCAAGCACTGCCCACTTTGGTACGGATACGGCGG 2784

```
Qy       2536 TGGGTTGAAATGGCTTGAGCGGTTGTCCTACATTAACTCTGTGGTTTACCCGTGGACCTC 2595
Db       2785 CGGGCTAAAATTCCTGGAAAGGTTTTCTTATATCAACTCCATCGTTTATCCCTGGACGTC 2844

Qy       2596 TCTACCGCTCATCGTTTACTGTTCTCTCCCTGCCATCTGTCTTCTCACTGGAAAATTCAT 2655
Db       2845 CATTCCTCTCCTGGCTTACTGTACCTTGCCTGCCATCTGCCTGCTCACGGGGAAGTTTAT 2904

Qy       2656 CGTTCCCGAGATTAGCAACTATGCGAGTATCCTCTTCATGGCGCTCTTCTCGTCGATTGC 2715
Db       2905 CACACCAGAGCTTACCAATGTCGCCAGTATCTGGTTCATGGCACTTTTCATCTGCATCTC 2964

Qy       2716 AATAACGGGTATTCTCGAGATGCAATGGGGCAAAGTTGGGATCGATGATTGGTGGAGAAA 2775
Db       2965 CGTGACCGGCATCCTGGAAATGAGGTGGAGTGGCGTGGCCATCGACGACTGGTGGAGGAA 3024

Qy       2776 CGAACAGTTTTGGGTCATTGGAGGTGTTTCTGCGCATCTGTTTGCTCTCTTCCAAGGTCT 2835
Db       3025 CGAGCAGTTCTGGGTCATCGGAGGCGTTTCGGCGCATCTGTTCGCGGTGTTCCAGGGCCT 3084

Qy       2836 CCTCAAGGTTCTTGCTGGTGTCGACACTAACTTCACAGTCACATCAAAAGCAGCTGATGA 2895
Db       3085 GCTGAAGGTGTTCGCCGGCATCGACACGAGCTTCACCGTGACGTCGAAGGCCGGGGACGA 3144

Qy       2896 TGGAGAGTTCTCTGACCTTTACCTCTTCAAATGGACTTCACTTCTCATCCCTCCAATGAC 2955
Db       3145 CGAGGAGTTCTCGGAGCTGTACACGTTCAAGTGGACCACCCTGCTGATACCCCCGACCAC 3204

Qy       2956 TCTACTCATCATAAACGTCATTGGAGTCATAGTCGGAGTCTCTGATGCCATCAGCAATGG 3015
Db       3205 GCTCCTCCTGCTGAACTTCATCGGGGTGGTGGCCGGGATCTCGAACGCGATCAACAACGG 3264

Qy       3016 ATACGACTCGTGGGGACCGCTTTTCGGAAGACTGTTCTTTGCACTTTGGGTCATCATTCA 3075
Db       3265 GTACGAGTCGTGGGGCCCCCTGTTCGGGAAGCTCTTCTTCGCCTTCTGGGTGATCGTCCA 3324

Qy       3076 TCTTTACCCGTTCCTTAAAGGTTTGCTTGGGAAACAAGATAGAATGCCAACCATTATTGT 3135
Db       3325 CCTGTACCCGTTCCTCAAGGGTCTGGTGGGGAGGCAGAACAGGACGCCGACGATCGTCAT 3384

Qy       3136 CGTCTGGTCCATCCTCCTGGCCTCGATTCTTACACTTCTTTGGGTCCGGGTTAATCCGTT 3195
Db       3385 CGTCTGGTCCATCCTGCTGGCCTCGATCTTCTCGCTCCTGTGGGTCCGCGTCGACCCGTT 3444

Qy       3196 TGTGGCGAAAGG---CGGTCCTATTCTCGAGATCTGTGGTTTAGACTGCTTGTGATTCGA 3252
Db       3445 CCTCGCCAAGAGCAACGGCCCGTTCCTGGAGGAGTGTGGCTTTAATTGCAACTGAAGTGG 3504

Qy       3253 TTGACCGGTG 3262
Db       3505 GGGCCCCCTG 3514
```

```
RESULT 15
US-10-160-719-45
; Sequence 45, Application US/10160719
; Publication No. US20030167528A1
; GENERAL INFORMATION:
;  APPLICANT: Kanwarpal S. Dhugga
;  APPLICANT:  Timothy G. Helentjaris
;  APPLICANT:  Benjamin A.  Bowen
;  APPLICANT:  Xun Wang
;  TITLE OF INVENTION: Maize Cellulose Synthases and Uses
;  TITLE OF INVENTION:  Thereof
;  FILE REFERENCE: 0864
;  CURRENT APPLICATION NUMBER: US/10/160,719
;  CURRENT FILING DATE:  2002-06-03
;  PRIOR APPLICATION NUMBER: US/09/371,383A
;  PRIOR FILING DATE: 1999-08-06
;  PRIOR APPLICATION NUMBER: US 60/096,822
;  PRIOR FILING DATE: 1998-08-17
;  NUMBER OF SEQ ID NOS: 60
;  SOFTWARE: FastSEQ for Windows Version 3.0
```

```
; SEQ ID NO 45
;   LENGTH: 3813
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (215)...(3496)
US-10-160-719-45

  Query Match            41.1%;  Score 1416.4;  DB 12;  Length 3813;
  Best Local Similarity  66.7%;  Pred. No. 0;
  Matches 2128;  Conservative    0;  Mismatches 1026;  Indels   36;  Gaps    6;

Qy        103 GACATGTCAAATCTGCAGAGATGAGATCGAATTGACTGTTGATGGAGAACCGTTTGTGGC 162
Db        331 GCCGTGCCAGATATGCGGCGACGAGGTCGGGGTGGGCTTCGACGGGGAGCCCTTCGTGGC 390

Qy        163 ATGTAACGAATGTGCATTCCCTGTGTGTAGACCTTGCTATGAGTACGAAAGACGAGAAGG 222
Db        391 GTGCAACGAGTGCGCCTTCCCCGTCTGCCGCGCCTGCTACGAGTACGAGCGCCGCGAGGG 450

Qy        223 CAATCAAGCTTGTCCACAGTGCAAAACCCGTTTCAAACGTCTTAAAGGAAGTCCAAGAGT 282
Db        451 CTCGCAAGCGTGCCCGCAGTGCAGGACCCGCTACAAGCGCCTCAAGGGCTGCCCGCGGGT 510

Qy        283 TGAAGGTGATGAAGAGGAAGATGACATTGATGATTTAGACAATGAGTT---------TGA 333
Db        511 GGCCGGCGACGAGGAGGAGGACGGCGTCGACGACCTGGAGGGCGAGTTCGGCCTGCAGGA 570

Qy        334 GTATGGAAATAATGGGATTGGATTTGATCAGGTTTCTGAAGGTATGTCAATCTCTCGTCG 393
Db        571 CGGCGCCGCCCACGAGGACGACCCGCAGTACGTCGCCGAGTCCATGCTCAGGGCGCAGAT 630

Qy        394 CAACTCCGGTTTCCCACAATCTGATTTGGATTCAGCTCCACCTGGCTCTCAGATTCCATT 453
Db        631 GAGCTACGGCCGCGGCGGCGACGCGCACCCCGGCTTCAGCCCCGTCCCCAACGTGCCGCT 690

Qy        454 GCTGACTTACGGCGACGAGGACGTTGAGATTTCTTCTGATAGACATGCTCTTATTGTTCC 513
Db        691 CCTCACCAACGGCCAGATGGTTGATGACATCCCGCCGGAGCAGCACGCGCTCGTGCCGTC 750

Qy        514 TCCTTCACTTGGTGGTCAT------GGCAATAGAGTTCATCCTGTTTCTCTTTCTGACCC 567
Db        751 CTACATGAGCGGCGGCGGCGGCGGGGGCAAGAGGATCCACCCGCTCCCTTTCGCAGATCC 810

Qy        568 GACCGTGGCTGCACATCGAAGGCTGATGGTACCTCAGAAAGATCTTGCGGTTTATGGTTA 627
Db        811 CAACCTTCCAGTGCAACCGAGATCCATGGACCCGTCCAAGGATCTGGCCGCCTACGGATA 870

Qy        628 TGGAAGTGTCGCTTGGAAAGATCGGATGGAGGAATGGAAGAGAAAGCAGAATGAGAAACT 687
Db        871 TGGCAGCGTGGCCTGGAAGGAGAGAATGGAGGGCTGGAAGCAGAAGCAG---GAGCGCCT 927

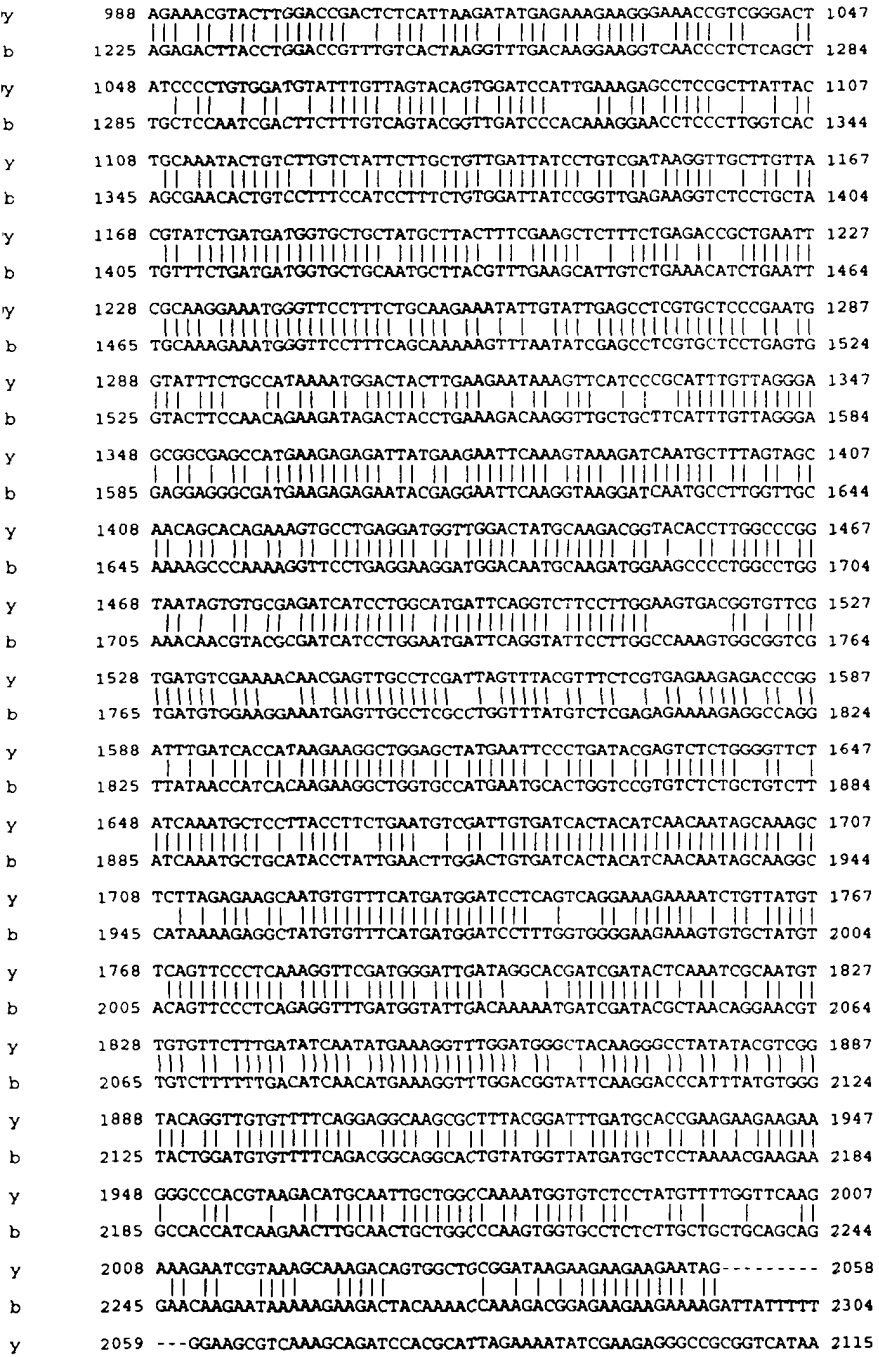
Qy        688 TCAGGTTGTTAGGCATGAAGGAGATCCTGATTTTGAAGATGGTGATGATGCTGATTTTCC 747
Db        928 GCAGCATGTCAGGAGCGAGGGTGGCGGTGATTGG---GATGGCGACGATGCAGATCTGCC 984

Qy        748 AATGATGGATGAGGGAAGGCAGCCATTGTCTATGAAGATACCAATCAAATCGAGCAAGAT 807
Db        985 ACTAATGGATGAAGCTAGGCAGCCATTGTCCAGAAAAGTCCCTATATCATCAAGCCGAAT 1044

Qy        808 AAATCCTTACCGGATGTTAATTGTGCTACGTCTTGTGATTCTTGGTCTCTTCTTTCACTA 867
Db       1045 TAATCCCTACAGGATGATTATCGTTATCCGGTTGGTGGTTTTGGGTTTCTTCTTCCACTA 1104

Qy        868 CCGTATTCTTCACCCCGTCAAAGATGCATATGCTTTGTGGCTTATTTCTGTTATATGTGA 927
Db       1105 CCGAGTGATGCATCCGGCGAAAGATGCATTTGCATTGTGGCTCATATCTGTAATCTGTGA 1164

Qy        928 GATATGGTTTGCTGTTTCATGGGTTCTTGATCAGTTCCCTAAATGGTACCCTATCGAGCG 987
Db       1165 AATCTGGTTTGCGATGTCCTGGATTCTTGATCAGTTCCCAAAGTGGCTTCCAATCGAGAG 1224
```


Qy 988 AGAAACGTACTTGGACCGACTCTCATTAAGATATGAGAAAGAAGGGAAACCGTCGGGACT 1047
Db 1225 AGAGACTTACCTGGACCGTTTGTCACTAAGGTTTGACAAGGAAGGTCAACCCTCTCAGCT 1284
Qy 1048 ATCCCCTGTGGATGTATTTGTTAGTACAGTGGATCCATTGAAAGAGCCTCCGCTTATTAC 1107
Db 1285 TGCTCCAATCGACTTCTTTGTCAGTACGGTTGATCCCACAAAGGAACCTCCCTTGGTCAC 1344
Qy 1108 TGCAAATACTGTCTTGTCTATTCTTGCTGTTGATTATCCTGTCGATAAGGTTGCTTGTTA 1167
Db 1345 AGCGAACACTGTCCTTTCCATCCTTTCTGTGGATTATCCGGTTGAGAAGGTCTCCTGCTA 1404
Qy 1168 CGTATCTGATGATGGTGCTGCTATGCTTACTTTCGAAGCTCTTTCTGAGACCGCTGAATT 1227
Db 1405 TGTTTCTGATGATGGTGCTGCAATGCTTACGTTTGAAGCATTGTCTGAAACATCTGAATT 1464
Qy 1228 CGCAAGGAAATGGGTTCCTTTCTGCAAGAAATATTGTATTGAGCCTCGTGCTCCCGAATG 1287
Db 1465 TGCAAAGAAATGGGTTCCTTTCAGCAAAAAGTTTAATATCGAGCCTCGTGCTCCTGAGTG 1524
Qy 1288 GTATTTCTGCCATAAAATGGACTACTTGAAGAATAAAGTTCATCCCGCATTTGTTAGGGA 1347
Db 1525 GTACTTCCAACAGAAGATAGACTACCTGAAAGACAAGGTTGCTGCTTCATTTGTTAGGGA 1584
Qy 1348 GCGGCGAGCCATGAAGAGAGATTATGAAGAATTCAAAGTAAAGATCAATGCTTTAGTAGC 1407
Db 1585 GAGGAGGGCGATGAAGAGAGAATACGAGGAATTCAAGGTAAGGATCAATGCCTTGGTTGC 1644
Qy 1408 AACAGCACAGAAAGTGCCTGAGGATGGTTGGACTATGCAAGACGGTACACCTTGGCCCGG 1467
Db 1645 AAAAGCCCAAAAGGTTCCTGAGGAAGGATGGACAATGCAAGATGGAAGCCCCTGGCCTGG 1704
Qy 1468 TAATAGTGTGCGAGATCATCCTGGCATGATTCAGGTCTTCCTTGGAAGTGACGGTGTTCG 1527
Db 1705 AAACAACGTACGCGATCATCCTGGAATGATTCAGGTATTCCTTGGCCAAAGTGGCGGTCG 1764
Qy 1528 TGATGTCGAAAACAACGAGTTGCCTCGATTAGTTTACGTTTCTCGTGAGAAGAGACCCGG 1587
Db 1765 TGATGTGGAAGGAAATGAGTTGCCTCGCCTGGTTTATGTCTCGAGAGAAAAGAGGCCAGG 1824
Qy 1588 ATTTGATCACCATAAGAAGGCTGGAGCTATGAATTCCCTGATACGAGTCTCTGGGGTTCT 1647
Db 1825 TTATAACCATCACAAGAAGGCTGGTGCCATGAATGCACTGGTCCGTGTCTCTGCTGTCTT 1884
Qy 1648 ATCAAATGCTCCTTACCTTCTGAATGTCGATTGTGATCACTACATCAACAATAGCAAAGC 1707
Db 1885 ATCAAATGCTGCATACCTATTGAACTTGGACTGTGATCACTACATCAACAATAGCAAGGC 1944
Qy 1708 TCTTAGAGAAGCAATGTGTTTCATGATGGATCCTCAGTCAGGAAAGAAAATCTGTTATGT 1767
Db 1945 CATAAAAGAGGCTATGTGTTTCATGATGGATCCTTTGGTGGGGAAGAAAGTGTGCTATGT 2004
Qy 1768 TCAGTTCCCTCAAAGGTTCGATGGGATTGATAGGCACGATCGATACTCAAATCGCAATGT 1827
Db 2005 ACAGTTCCCTCAGAGGTTTGATGGTATTGACAAAAATGATCGATACGCTAACAGGAACGT 2064
Qy 1828 TGTGTTCTTTGATATCAATATGAAAGGTTTGGATGGGCTACAAGGGCCTATATACGTCGG 1887
Db 2065 TGTCTTTTTTGACATCAACATGAAAGGTTTGGACGGTATTCAAGGACCCATTTATGTGGG 2124
Qy 1888 TACAGGTTGTGTTTTCAGGAGGCAAGCGCTTTACGGATTTGATGCACCGAAGAAGAAGAA 1947
Db 2125 TACTGGATGTGTTTTCAGACGGCAGGCACTGTATGGTTATGATGCTCCTAAAACGAAGAA 2184
Qy 1948 GGGCCCACGTAAGACATGCAATTGCTGGCCAAAATGGTGTCTCCTATGTTTTGGTTCAAG 2007
Db 2185 GCCACCATCAAGAACTTGCAACTGCTGGCCCAAGTGGTGCCTCTCTTGCTGCTGCAGCAG 2244
Qy 2008 AAAGAATCGTAAAGCAAAGACAGTGGCTGCGGATAAGAAGAAGAAGAATAG--------- 2058
Db 2245 GAACAAGAATAAAAAGAAGACTACAAAACCAAAGACGGAGAAGAAGAAAAGATTATTTTT 2304
Qy 2059 ---GGAAGCGTCAAAGCAGATCCACGCATTAGAAAATATCGAAGAGGGCCGCGGTCATAA 2115
Db 2305 CAAGAAAGCAGAAAACCCATCTCCTGCATATGCTTTGGGTGAAATTGATGAAGGTGCTCC 2364
Qy 2116 AGTTCTTAACGTAGAACAGTCAACCGAGGCAATGCAAATGAAGTTGCAGAAGAAATATGG 2175
Db 2365 AGGTGCTGATATCGAGAAGGCCGGAATCGTAAATCAACAGAAACTAGAGAAGAAATTTGG 2424
Qy 2176 GCAGTCTCCTGTATTTGTTGCATCTGCGCGTCTGGAGAATGGTGGGATGGCTAGAAACGC 2235
Db 2425 GCAGTCTTCTGTTTTTGTCGCATCAACACTTCTTGAGAACGGAGGGACCCTGAAGAGCGC 2484
Qy 2236 AAGCCCGGCTTGTCTGCTTAAAGAAGCCATCCAAGTCATTAGTCGCGGATATGAAGATAA 2295
Db 2485 AAGTCCAGCTTCTCTTCTGAAGGAAGCTATACATGTTATCAGCTGCGGCTACGAAGACAA 2544
Qy 2296 AACTGAATGGGGAAAAGAGATTGGGTGGATCTATGGTTCTGTTACCGAAGATATTCTTAC 2355
Db 2545 GACCGACTGGGGAAAAGAGATTGGCTGGATTTACGGATCGATCACAGAGGATATCTTGAC 2604
Qy 2356 GGGTTCTAAGATGTATTCTCATGGTTGGAGACATGTTTATTGTACACCAAAGTTAGCGGC 2415
Db 2605 TGGATTTAAGATGTACTGCCATGGCTGGGGTCTATTTACTGCATCCCGAAGCGGCCTGC 2664
Qy 2416 TTTCAAAGGATCAGCTCCAATCAATCTTTCGGATCGTCTCCATCAAGTTCTTCGATGGGC 2475
Db 2665 ATTCAAAGGTTCTGCGCCTCTGAACCTTTCCGACCGTCTTCACCAGGTCCTTCGCTGGGC 2724
Qy 2476 GCTTGGGTCGGTTGAGATTTTCTTGAGTAGGCATTGTCCTATTTGGTATGGTTATGGAGG 2535
Db 2725 CCTTGGGTCCGTCGAAATTTTCTTCAGCAAGCACTGCCCACTTTGGTACGGATACGGCGG 2784
Qy 2536 TGGGTTGAAATGGCTTGAGCGGTTGTCCTACATTAACTCTGTGGTTTACCCGTGGACCTC 2595
Db 2785 CGGGCTAAAATTCCTGGAAAGGTTTTCTTATATCAACTCCATCGTTTATCCCTGGACGTC 2844
Qy 2596 TCTACCGCTCATCGTTTACTGTTCTCTCCCTGCCATCTGTCTTCTCACTGGAAAATTCAT 2655
Db 2845 CATTCCTCTCCTGGCTTACTGTACCTTGCCTGCCATCTGCCTGCTCACGGGGAAGTTTAT 2904
Qy 2656 CGTTCCCGAGATTAGCAACTATGCGAGTATCCTCTTCATGGCGCTCTTCTCGTCGATTGC 2715
Db 2905 CACACCAGAGCTTACCAATGTCGCCAGTATCTGGTTCATGGCACTTTTCATCTGCATCTC 2964
Qy 2716 AATAACGGGTATTCTCGAGATGCAATGGGGCAAAGTTGGGATCGATGATTGGTGGAGAAA 2775
Db 2965 CGTGACCGGCATCCTGGAAATGAGGTGGAGTGGCGTGGCCATCGACGACTGGTGGAGGAA 3024
Qy 2776 CGAACAGTTTTGGGTCATTGGAGGTGTTTCTGCGCATCTGTTTGCTCTCTTCCAAGGTCT 2835
Db 3025 CGAGCAGTTCTGGGTCATCGGAGGCGTTTCGGCGCATCTGTTCGCGGTGTTCCAGGGCCT 3084
Qy 2836 CCTCAAGGTTCTTGCTGGTGTCGACACTAACTTCACAGTCACATCAAAAGCAGCTGATGA 2895
Db 3085 GCTGAAGGTGTTCGCCGGCATCGACACGAGCTTCACCGTGACGTCGAAGGCCGGGGACGA 3144
Qy 2896 TGGAGAGTTCTCTGACCTTTACCTCTTCAAATGGACTTCACTTCTCATCCCTCCAATGAC 2955
Db 3145 CGAGGAGTTCTCGGAGCTGTACACGTTCAAGTGGACCACCCTGCTGATACCCCCGACCAC 3204
Qy 2956 TCTACTCATCATAAACGTCATTGGAGTCATAGTCGGAGTCTCTGATGCCATCAGCAATGG 3015
Db 3205 GCTCCTCCTGCTGAACTTCATCGGGGTGGTGGCCGGGATCTCGAACGCGATCAACAACGG 3264
Qy 3016 ATACGACTCGTGGGGACCGCTTTTCGGAAGACTGTTCTTTGCACTTTGGGTCATCATTCA 3075
Db 3265 GTACGAGTCGTGGGGCCCCCTGTTCGGGAAGCTCTTCTTCGCCTTCTGGGTGATCGTCCA 3324
Qy 3076 TCTTTACCCGTTCCTTAAAGGTTTGCTTGGGAAACAAGATAGAATGCCAACCATTATTGT 3135
Db 3325 CCTGTACCCGTTCCTCAAGGGTCTGGTGGGGAGGCAGAACAGGACGCCGACGATCGTCAT 3384
Qy 3136 CGTCTGGTCCATCCTCCTGGCCTCGATTCTTACACTTCTTTGGGTCCGGGTTAATCCGTT 3195

```
Db       3385 CGTCTGGTCCATCCTGCTGGCCTCGATCTTCTCGCTCCTGTGGGTCCGCGTCGACCCGTT 3444

Qy       3196 TGTGGCGAAAGG---CGGTCCTATTCTCGAGATCTGTGGTTTAGACTGCTTGTGATTCGA 3252
                | || ||  |   ||| ||  | || |||   ||||  | ||||||   |||   |
Db       3445 CCTCGCCAAGAGCAACGGCCCGCTCCTGGAGGAGTGTGGCCTGGACTGCAACTGAAGTGG 3504

Qy       3253 TTGACCGGTG 3262
                | ||  ||
Db       3505 GGGCCCCCTG 3514


Search completed: October 30, 2003, 22:43:32
Job time : 875.127 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic nucleic search, using sw model

Run on: October 30, 2003, 05:38:04 ; Search time 181.589 Seconds
(without alignments)
8371.232 Million cell updates/sec

Title: US-09-980-043-5
Perfect score: 3444
Sequence: 1 gcggccgcggttaatcgccg.....gtgttgtagatagcggccgc 3444

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1934.4	56.2	3828	4	US-09-221-013A-7	Sequence 7, Appli
2	1294.8	37.6	3603	4	US-09-221-013A-5	Sequence 5, Appli
3	1294.8	37.6	3673	4	US-09-221-013A-11	Sequence 11, Appl
4	1252.6	36.4	3614	4	US-09-221-013A-9	Sequence 9, Appli
5	1145	33.2	3328	3	US-08-960-048-1	Sequence 1, Appli
6	1145	33.2	3328	4	US-09-838-586-1	Sequence 1, Appli
7	822.2	23.9	2248	4	US-09-221-013A-1	Sequence 1, Appli
8	546.2	15.9	1741	4	US-09-221-013A-13	Sequence 13, Appl
9	484.2	14.1	5009	4	US-09-221-013A-4	Sequence 4, Appli
10	375	10.9	8411	4	US-09-221-013A-3	Sequence 3, Appli
11	155.8	4.5	281	4	US-09-313-294A-6485	Sequence 6485, Ap
12	137.8	4.0	291	4	US-09-313-294A-5519	Sequence 5519, Ap
13	137.4	4.0	268	4	US-09-313-294A-849	Sequence 849, App
14	132.4	3.8	300	4	US-09-313-294A-4753	Sequence 4753, Ap
15	123	3.6	286	4	US-09-313-294A-4696	Sequence 4696, Ap
16	101.4	2.9	278	4	US-09-313-294A-4015	Sequence 4015, Ap
17	99	2.9	317	4	US-09-313-294A-7123	Sequence 7123, Ap
18	96.4	2.8	283	4	US-09-313-294A-1981	Sequence 1981, Ap
19	85.4	2.5	287	4	US-09-313-294A-5274	Sequence 5274, Ap
20	80.2	2.3	159	4	US-09-313-294A-1693	Sequence 1693, Ap
21	69.4	2.0	284	4	US-09-313-294A-3993	Sequence 3993, Ap
22	63.2	1.8	250	4	US-09-313-294A-582	Sequence 582, App
c 23	61.6	1.8	7218	1	US-08-232-463-14	Sequence 14, Appl
24	52.8	1.5	4612	3	US-08-960-048-2	Sequence 2, Appli
25	52.8	1.5	4612	4	US-09-838-586-2	Sequence 2, Appli
26	45.6	1.3	7218	1	US-08-232-463-14	Sequence 14, Appl
c 27	41.4	1.2	6794	4	US-09-491-356C-2	Sequence 2, Appli
28	41	1.2	1437	4	US-09-137-223A-4	Sequence 4, Appli
c 29	37.6	1.1	19877	2	US-08-816-155B-8	Sequence 8, Appli
c 30	37.6	1.1	19877	3	US-09-079-587-8	Sequence 8, Appli
c 31	37.4	1.1	265	4	US-09-491-356C-14	Sequence 14, Appl
c 32	37.4	1.1	265	4	US-09-491-356C-16	Sequence 16, Appl
c 33	37.4	1.1	265	4	US-09-491-356C-17	Sequence 17, Appl
c 34	37.4	1.1	265	4	US-09-491-356C-18	Sequence 18, Appl
c 35	37	1.1	87350	3	US-08-781-891-79	Sequence 79, Appl
c 36	37	1.1	87350	4	US-09-618-166-79	Sequence 79, Appl
c 37	37	1.1	87543	4	US-09-791-211-3	Sequence 3, Appli
c 38	36.8	1.1	486	4	US-09-522-217-3	Sequence 3, Appli
39	36.4	1.1	2518	3	US-09-433-699-3	Sequence 3, Appli
c 40	36	1.0	2190	4	US-09-625-188-19	Sequence 19, Appl
c 41	35.8	1.0	253	4	US-09-491-356C-13	Sequence 13, Appl
c 42	35.8	1.0	55298	4	US-09-491-356C-1	Sequence 1, Appli
43	35.6	1.0	2295	1	US-08-375-300-3	Sequence 3, Appli
44	35.6	1.0	2295	3	US-09-177-431-3	Sequence 3, Appli
45	35.6	1.0	2295	5	PCT-US95-16930-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-09-221-013A-7
; Sequence 7, Application US/09221013A
; Patent No. 6495740
; GENERAL INFORMATION:
; APPLICANT: Arioli, Antonio
; APPLICANT: Williamson, Richard E.
; APPLICANT: Betzner, Andreas S.
; APPLICANT: Peng, Liangcai
; TITLE OF INVENTION: Manipulation of cellulose and/or Beta-1,4-glucan
; FILE REFERENCE: 96-98
; CURRENT APPLICATION NUMBER: US/09/221,013A
; CURRENT FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: PCT/AU97/00402
; PRIOR FILING DATE: 1997-06-24
; PRIOR APPLICATION NUMBER: AU PO0699
; PRIOR FILING DATE: 1996-06-27
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 3828
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (239)..(3490)
US-09-221-013A-7

Query Match 56.2%; Score 1934.4; DB 4; Length 3828;
Best Local Similarity 76.2%; Pred. No. 0;
Matches 2478; Conservative 0; Mismatches 731; Indels 45; Gaps 6;

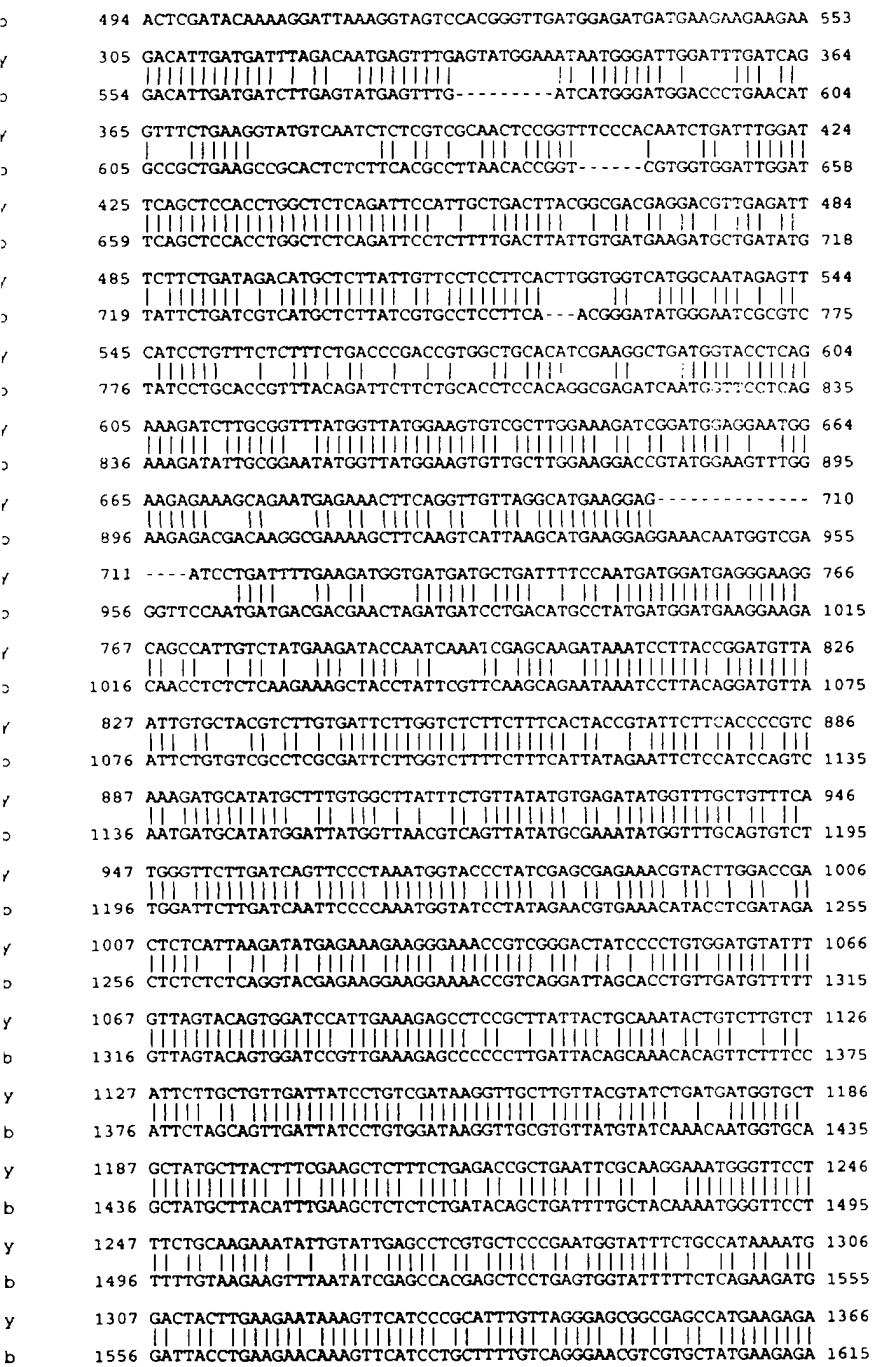
```
Qy          8 CGGTTAATCGCCGGTTCTCACAACAGGAATGAGTTTGTCCTCATTAATGCCGATGAGAAT 67
              ||| | || || || |||||||||| || || || || |||||||| |||||||||| |
Db        254 CGGCTCATTGCTGGCTCTCACAACAGAAACGAATTCGTTCTCATTAACGCCGATGAGAGT 313

Qy         68 GCCCGAATAAGATCAGTCCAAGAGCTGAGTGGACAGACATGTCAAATCTGCAGAGATGAG 127
              ||| ||||| ||||||| |||| |||||||| || ||||||||||||||  |||||||
Db        314 GCCAGAATACGATCAGTACAAGAACTGAGTGGGCAAACATGTCAAATCTGTGGAGATGAA 373

Qy        128 ATCGAATTGACTGTTGATGGAGAACCGTTTGTGGCATGTAACGAATGTGCATTCCCTGTG 187
              |||||||| || |||     | || |  ||||| || || |||||||| |||||||| ||
Db        374 ATCGAATTAACGGTTAGCAGTGAGCTCTTTGTTGCTTGCAACGAATGCGCATTCCCGGTT 433

Qy        188 TGTAGACCTTGCTATGAGTACGAAAGACGAGAAGGCAATCAAGCTTGTCCACAGTGCAAA 247
              |||||||| |||||||||| ||| |  ||||||| |||||||||||||| |||||||||
Db        434 TGTAGACCATGCTATGAGTATGAACGTAGAGAAGGAAATCAAGCTTGTCCTCAGTGCAAA 493

Qy        248 ACCCGTTTCAAACGTCTTAAAGGAAGTCCAAGAGTTGA---AGGTGATGAAGAGGAAGAT 304
              || || | |||| |  | ||||| |||||| | |||||   || |||||||||| |||||
```


Db 494 ACTCGATACAAAAGGATTAAAGGTAGTCCACGGGTTGATGGAGATGATGAAGAAGAAGAA 553
Qy 305 GACATTGATGATTTAGACAATGAGTTTGAGTATGGAAATAATGGGATTGGATTTGATCAG 364
Db 554 GACATTGATGATCTTGAGTATGAGTTTG---------ATCATGGGATGGACCCTGAACAT 604
Qy 365 GTTTCTGAAGGTATGTCAATCTCTCGTCGCAACTCCGGTTTCCCACAATCTGATTTGGAT 424
Db 605 GCCGCTGAAGCCGCACTCTCTTCACGCCTTAACACCGGT------CGTGGTGGATTGGAT 658
Qy 425 TCAGCTCCACCTGGCTCTCAGATTCCATTGCTGACTTACGGCGACGAGGACGTTGAGATT 484
Db 659 TCAGCTCCACCTGGCTCTCAGATTCCTCTTTTGACTTATTGTGATGAAGATGCTGATATG 718
Qy 485 TCTTCTGATAGACATGCTCTTATTGTTCCTCCTTCACTTGGTGGTCATGGCAATAGAGTT 544
Db 719 TATTCTGATCGTCATGCTCTTATCGTGCCTCCTTCA---ACGGGATATGGGAATCGCGTC 775
Qy 545 CATCCTGTTTCTCTTTCTGACCCGACCGTGGCTGCACATCGAAGGCTGATGGTACCTCAG 604
Db 776 TATCCTGCACCGTTTACAGATTCTTCTGCACCTCCACAGGCGAGATCAATGGTTCCTCAG 835
Qy 605 AAAGATCTTGCGGTTTATGGTTATGGAAGTGTCGCTTGGAAAGATCGGATGGAGGAATGG 664
Db 836 AAAGATATTGCGGAATATGGTTATGGAAGTGTTGCTTGGAAGGACCGTATGGAAGTTTGG 895
Qy 665 AAGAGAAAGCAGAATGAGAAACTTCAGGTTGTTAGGCATGAAGGAG-------------- 710
Db 896 AAGAGACGACAAGGCGAAAAGCTTCAAGTCATTAAGCATGAAGGAGGAAACAATGGTCGA 955
Qy 711 ----ATCCTGATTTTGAAGATGGTGATGATGCTGATTTTCCAATGATGGATGAGGGAAGG 766
Db 956 GGTTCCAATGATGACGACGAACTAGATGATCCTGACATGCCTATGATGGATGAAGGAAGA 1015
Qy 767 CAGCCATTGTCTATGAAGATACCAATCAAATCGAGCAAGATAAATCCTTACCGGATGTTA 826
Db 1016 CAACCTCTCTCAAGAAAGCTACCTATTCGTTCAAGCAGAATAAATCCTTACAGGATGTTA 1075
Qy 827 ATTGTGCTACGTCTTGTGATTCTTGGTCTCTTCTTTCACTACCGTATTCTTCACCCCGTC 886
Db 1076 ATTCTGTGTCGCCTCGCGATTCTTGGTCTTTTCTTTCATTATAGAATTCTCCATCCAGTC 1135
Qy 887 AAAGATGCATATGCTTTGTGGCTTATTTCTGTTATATGTGAGATATGGTTTGCTGTTTCA 946
Db 1136 AATGATGCATATGGATTATGGTTAACGTCAGTTATATGCGAAATATGGTTTGCAGTGTCT 1195
Qy 947 TGGGTTCTTGATCAGTTCCCTAAATGGTACCCTATCGAGCGAGAAACGTACTTGGACCGA 1006
Db 1196 TGGATTCTTGATCAATTCCCCAAATGGTATCCTATAGAACGTGAAACATACCTCGATAGA 1255
Qy 1007 CTCTCATTAAGATATGAGAAAGAAGGGAAACCGTCGGGACTATCCCCTGTGGATGTATTT 1066
Db 1256 CTCTCTCTCAGGTACGAGAAGGAAGGAAAACCGTCAGGATTAGCACCTGTTGATGTTTTT 1315
Qy 1067 GTTAGTACAGTGGATCCATTGAAAGAGCCTCCGCTTATTACTGCAAATACTGTCTTGTCT 1126
Db 1316 GTTAGTACAGTGGATCCGTTGAAAGAGCCCCCCTTGATTACAGCAAACACAGTTCTTTCC 1375
Qy 1127 ATTCTTGCTGTTGATTATCCTGTCGATAAGGTTGCTTGTTACGTATCTGATGATGGTGCT 1186
Db 1376 ATTCTAGCAGTTGATTATCCTGTGGATAAGGTTGCGTGTTATGTATCAAACAATGGTGCA 1435
Qy 1187 GCTATGCTTACTTTCGAAGCTCTTTCTGAGACCGCTGAATTCGCAAGGAAATGGGTTCCT 1246
Db 1436 GCTATGCTTACATTTGAAGCTCTCTCTGATACAGCTGATTTTGCTACAAAATGGGTTCCT 1495
Qy 1247 TTCTGCAAGAAATATTGTATTGAGCCTCGTGCTCCCGAATGGTATTTCTGCCATAAAATG 1306
Db 1496 TTTTGTAAGAAGTTTAATATCGAGCCACGAGCTCCTGAGTGGTATTTTTCTCAGAAGATG 1555
Qy 1307 GACTACTTGAAGAATAAAGTTCATCCCGCATTTGTTAGGGAGCGGCGAGCCATGAAGAGA 1366
Db 1556 GATTACCTGAAGAACAAAGTTCATCCTGCTTTTGTCAGGGAACGTCGTGCTATGAAGAGA 1615
Qy 1367 GATTATGAAGAATTAAAGTAAAGATCAATGCTTTAGTAGCAACAGCACAGAAAGTGCCT 1426
Db 1616 GATTATGAAGAGTTTAAAGTGAAGATAAATGCACTGGTTGCTACTGCACAGAAAGTGCCT 1675
Qy 1427 GAGGATGGTTGGACTATGCAAGACGGTACACCTTGGCCCGGTAATAGTGTGCGAGATCAT 1486
Db 1676 GAGGAACGTTGGACTATGCAAGATGGAACTCCTTGGCCTGGAAACAACGTCCGTGACCAT 1735
Qy 1487 CCTGGCATGATTCAAGTCTTCCTTGGAAGTGACGGTGTTCGTGATGTCGAAAACAACGAG 1546
Db 1736 CCTGGAATGATTCAAGTGTTCTTGGGTCATAGTGGAGTTCGTGATACGGATGGTAATGAG 1795
Qy 1547 TTGCCTCGATTAGTTTACGTTTCTCGTGAGAAGAGACCCGGATTTGATCACCATAAGAAG 1606
Db 1796 TTACCACGTCTAGTGTATGTTTCTCGTGAGAAGCGGCCTGGATTTGATCACCACAAGAAA 1855
Qy 1607 GCTGGAGCTATGAAGTCGCTGATACGAGTCTCGTGGGGTTCTATCAAATGCTCCTTACCTT 1666
Db 1856 GCTGGAGCTATGAAGTCGCTTGATACGAGTCTCGTGGTTCTATCAAACGCTCCTTACCTT 1915
Qy 1667 CTGAATGTCGATTGTGATCACTACATCAACAACAGCAAAGCTCTTAGAGAAGCAATGTGT 1726
Db 1916 CTTAATGTCGATTGTGATCACTACATCAACAACAGCAAAGCAATTAGAGAATCTATGTGT 1975
Qy 1727 TTCATGATGGATCCTCAGTCAGGAAAGAAAATCTGTTATGTTCAGTTCCCTCAAAGGTTC 1786
Db 1976 TTCATGATGGACCCGCAATCGGGAAAGAAAGTTGTTATGTTCAGTTTCCGCAGAGATTT 2035
Qy 1787 GATGGGATTGATAGGCACGATCGATACTCAAATCGCAATGTTGTGTTCTTTGATATCAAT 1846
Db 2036 GATGGGATTGATAGACATGATAGATACTCAAACCGTAACGTTGTGTTCTTTGATATTAAC 2095
Qy 1847 ATGAAAGGTTTGGATGGGCTACAAGGGCCTATATACGTCGGTACAGGTTGTGTTTTCAGG 1906
Db 2096 ATGAAAGGTCTTGATGGGATACAAGGACCGATATATGTCGGGACAGGTTGTGTGTTTAGA 2155
Qy 1907 AGGCAAGCGCTTTACGGATTTGATGCACCGAAGAAGAAGAAGGGCCCACGTAAGACATGC 1966
Db 2156 AAACAGGCTCTTTATGGTTTTGATGCACCAAAGAAGAAGAAACCACCAGGCAAAACCTGT 2215
Qy 1967 AATTGCTGGCCAAAATGGTGTCTCCTATGTTTTGGTTCAAGAAAGAATCGTAAAGCAAAG 2026
Db 2216 AACTGTTGGCCTAAATGGTGTTGTTTGTGTTGTGGGTTGAGAAAGAAGAGTAAAACGAAA 2275
Qy 2027 ACAGTGGCTGCGGATAAGAAGAAGAAGAATAGGGAAGCGTCAAAGCAGATCCACGCATTA 2086
Db 2276 ------GCCACAGATAAGAAAACTAACACTAAAGAGACTTCAAAGCAGATTCATGCGCTA 2329
Qy 2087 GAAAATATCGAAGAGGGCCGCGGTCATAAAGTTCTTAACGTAGAACAGTCAACCGAGGCA 2146
Db 2330 GAGAATGTCGACGAAGGTGTTATCGTCCCAGTGTCAAATGTTGAGAAGAGATCTGAAGCA 2389
Qy 2147 ATGCAAATGAAGTTGCAGAAGAAATATGGGCAGTCTCCTGTATTTGTTGCATCTGCGCGT 2206
Db 2390 ACACAATTGAAATTGGAGAAGAAGTTTGGACAATCTCCGGTTTTCGTTGCCTCTGCTGTT 2449
Qy 2207 CTGGAGAATGGTGGGATGGCTAGAAACGCAAGCCCGGCTTGTCTGCTTAAAGAAGCCATC 2266
Db 2450 CTACAGAACGGTGGAGTTCCCCGTAACGCAAGCCCCGCATGTTTGTTAAGAGAAGCCATT 2509
Qy 2267 CAAGTCATTAGTCGCGGATATGAAGATAAAACTGAATGGGGAAAAGAGATTGGGTGGATC 2326
Db 2510 CAAGTTATTAGCTGCGGGTACCAAGATAAAACCGAATGGGGAAAAGAGATCGGGTGGATT 2569
Qy 2327 TATGGTTCTGTTACCGAAGATATTCTTACGGGTTCTAAGATGCATTCTCATGGTTGGAGA 2386
Db 2570 TATGGATCGGTGACTGAAGATATCCTGACGGGTTTCAAGATGCATTGCCATGGATGGAGA 2629
Qy 2387 CATGTTTATTGTACACCAAAGTTAGCGGCTTTCAAAGGATCAGCTCCAATCAATCTTTCG 2446
Db 2630 TCTGTGTACTGTATGCCTAAGCGTGCAGCTTTTAAAGGATCTGCTCCTATTAACTTGTCA 2689

```
Qy       2447 GATCGTCTCCATCAAGTTCTTCGATGGGCGCTTGGGTCGGTTGAGATTTTCTTGAGTAGG 2506
             |||||||| |||||||||| || |||| |||| || || ||||||||||||| ||
Db       2690 GATCGTCTTCATCAAGTTCTACGTTGGGCTCTTGGCTCTGTAGAGATTTTCTTGAGCAGA 2749

Qy       2507 CATTGTCCTATTTGGTATGGTTATGGAGGTGGGTTGAAATGGCTTGAGCGGTTGTCCTAC 2566
             |||||||| || |||||||||||||| ||||| || |||||| | ||| | || || |||
Db       2750 CATTGTCCGATATGGTATGGTTATGGTGGTGGTTTAAAATGGTTGGAGAGATTCTCTTAC 2809

Qy       2567 ATTAACTCTGTGGTTTACCCGTGGACCTCTCTACCGCTCATCGTTTACTGTTCTCTCCCT 2626
             || |||||||| || || || |||| || || ||  | |||| || ||||||||||
Db       2810 ATCAACTCTGTCGTCTATCCTTGGACTTCACTTCCATTGATCGTCTATTGTTCTCTCCCC 2869

Qy       2627 GCCATCTGTCTTCTCACTGGAAAATTCATCGTTCCCGAGATTAGCAACTATGCGAGTATC 2686
             ||  | ||| | |||| |||||||||||||| || |||| |||||||| ||  ||||
Db       2870 GCGGTTTGTTTACTCACAGGAAAATTCATCGTCCCTGAGATAAGCAACTACGCAGGTATA 2929

Qy       2687 CTCTTCATGGCGCTCTTCTCGTCGATTGCAATAACGGGTATTCTCGAGATGCAATGGGGC 2746
             |||||||||    | |||   || || ||| |||| || || ||||| ||||||||||
Db       2930 CTCTTCATGCTCATGTTCATATCCATAGCAGTAACTGGAATCCTCGAAATGCAATGGGGA 2989

Qy       2747 AAAGTTGGGATCGATGATTGGTGGAGAAACGAACAGTTTTGGGTCATTGGAGGTGTTTCT 2806
                || || ||||||||||||||||||||||||| |||||||||| || ||||| |  ||
Db       2990 GGTGTCGGAATCGATGATTGGTGGAGAAACGAGCAGTTTTGGGTAATCGGAGGGGCCTCC 3049

Qy       2807 GCGCATCTGTTTGCTCTCTTCCAAGGTCTCCTCAAGGTTCTTGCTGGTGTCGACACTAAC 2866
              ||||||| |||||||| || |||||| | ||||| |||| || || ||   |||| ||
Db       3050 TCGCATCTATTTGCTCTGTTTCAAGGTTTGCTCAAAGTTCTAGCCGGAGTTAACACGAAT 3109

Qy       2867 TTCACAGTCACATCAAAAGCAGCTGATGATGGAGAGTTCTCTGACCTTTACCTCTTCAAA 2926
             |||||||||| |||||||||| || |||||||  |||||||| |||||| |||||||
Db       3110 TTCACAGTCACTTCAAAAGCAGCAGACGATGGAGCTTTCTCTGAGCTTTACATCTTCAAG 3169

Qy       2927 TGGACTTCACTTCTCATCCCTCCAATGACTCTACTCATCATAAACGTCATTGGAGTCATA 2986
             |||||  |  |  | || ||||| |  || || || ||||| ||| ||||||||| ||
Db       3170 TGGACAACTTTGTTGATTCCTCCGACAACACTTCTGATCATTAACATCATTGGAGTTATT 3229

Qy       2987 GTCGGAGTCTCTGATGCCATCAGCAATGGATACGACTCGTGGGGACCGCTTTTCGGAAGA 3046
             ||||| || |||||||||| |||||||| || ||||| |||||||| || || || |||
Db       3230 GTCGGCGTTTCTGATGCCATTAGCAATGGCTATGACTCATGGGGACCTCTCTTTGGGAGA 3289

Qy       3047 CTGTTCTTTGCACTTTGGGTCATCATTCATCTTTACCCGTTCCTTAAAGGTTTGCTTGGG 3106
             || ||||| || |||||||||| |||| | |||| |||| || ||  ||||||||
Db       3290 CTTTTCTTCGCTCTTTGGGTCATTGTTCATTTATACCCATTCCTCAAGGGAATGCTTGGG 3349

Qy       3107 AAACAAGATAGAATGCCAACCATTATTGTCGTCTGGTCCATCCTCCTGGCCTCGATTCTT 3166
             || ||||| | |||||| || ||||| |||||||| || || || |||||  |
Db       3350 AAGCAAGACAAAATGCCTACGATTATTGTGGTCTGGTCTATTCTTCTAGCTTCGATCTTG 3409

Qy       3167 ACACTTCTTTGGGTCCGGGTTAATCCGTTTGTGGCGAAAGG GGTCCTATTCTCGAGATC 3226
             |||||  | ||||||  |  |||| |||||||||| |||| || ||  |  | ||||||
Db       3410 ACACTCTTGTGGGTCAGAATTAACCCGTTTGTGGCTAAAGG GGACCAGTGTTGGAGATC 3469

Qy       3227 TGTGGTTTAGACTG 3240
             |||||| |  | ||
Db       3470 TGTGGTCTGAATTG 3483


RESULT 2
US-09-221-013A-5
; Sequence 5, Application US/09221013A
; Patent No. 6495740
; GENERAL INFORMATION:
;  APPLICANT:  Arioli, Antonio
;  APPLICANT:  Williamson, Richard E.
;  APPLICANT:  Betzner, Andreas S.
;  APPLICANT:  Peng, Liangcai
;  TITLE OF INVENTION:  Manipulation of cellulose and/or Beta-1,4-glucan
;  FILE REFERENCE:  96-98
;  CURRENT APPLICATION NUMBER: US/09/221,013A
;  CURRENT FILING DATE:  1998-12-23
;  PRIOR APPLICATION NUMBER:  PCT/AU97/00402
;  PRIOR FILING DATE:  1997-06-24
;  PRIOR APPLICATION NUMBER:  AU PO0699
;  PRIOR FILING DATE:  1996-06-27
;  NUMBER OF SEQ ID NOS:  37
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 5
;   LENGTH: 3603
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(3243)
US-09-221-013A-5

  Query Match             37.6%;  Score 1294.8;  DB 4;  Length 3603;
  Best Local Similarity   64.5%;  Pred. No. 0;
  Matches 2074;  Conservative    0;  Mismatches 1082;  Indels   60;  Gaps    7;

Qy          9 GGTTAATCGCCGGTTCTCACAACAGGAATGAGTTTGTCCTCATTAATGCCGATGAGAATG 68
              | ||  | ||  ||  ||  ||   ||  || ||| | ||  |  ||        |||
Db         17 GCTTGGTTGCTGGATCCTACCGGAGAAACGAGCTCGTTCGGATCCGACATGAATCTGATG 76

Qy         69 CCCGAATAAGATCAGTCCAAGAGCTGAGTGGACAGACATGTCAAATCTGCAGAGATGAGA 128
              | | |  | | |   |  |  |  ||| ||| |||| |||||| |||||  | |||||
Db         77 GCGGGACCAAACCTTTGAAGAATATGAATGGCCAGATATGTCAGATCTGTGGTGATGATG 136

Qy        129 TCGAATTGACTGTTGATGGAGAACCGTTTGTGGCATGTAACGAATGTGCATTCCCTGTGT 188
              | | | |  |||     ||||||     |||||  || ||||| ||||||||  ||||||||||
Db        137 TTGGACTCGCTGAAACTGGAGATGTCTTTGTCGCGTGTAATGAATGTGCCTTCCCTGTGT 196

Qy        189 GTAGACCTTGCTATGAGTACGAAAGACGAGAAGGCAATCAAGCTTGTCCACAGTGCAAAA 248
              || | |||||||||||||||||||  |||  || | |||    ||| || | ||||| |
Db        197 GTCGGCCTTGCTATGAGTACGAGAGGAAAGATGGAACTCAGTGTTGCCCTCAATGCAAGA 256

Qy        249 CCCGTTTCAAACGTCTTAAAGGAAGTCCAAGAGTTGAAGGTGATGAAGAGGAAGATGACA 308
              |  | |||| ||| |  |  || |||||  | |||||||| |||||||| || |||||
Db        257 CTAGATTCAGACGACACAGGGGGAGTCCTCGTGTTGAAGGAGATGAAGATGAGGATGATG 316

Qy        309 TTGATGATTTAGACAATGAGTTTGAGTATGGAAATAATGGGATTGGATTTGATCAGGTTT 368
              |||||||| | || |||||||||  | ||                          |  ||||
Db        317 TTGATGATATCGAGAATGAGTTCAATTA----------------------CGCCCAGGGAG 355

Qy        369 CTGAAGGTATGTCAATCTCTCGTCGCAACTCCGGTTTCCCACAATCTGATTTGGATTCAG 428
              || |      |  |  |  |  ||  |    |  |  |  ||  |    ||  |
Db        356 CTAACAAGGCGAGACACCAACGCCATGGCGAAGAGTTTTCTTCTTCCTCTAGACATGAAT 415

Qy        429 CTCCACCTGGCTCTCAGATTCCATTGCTGACTTACGGCGACGAGGACGTTGAGATTTCTT 488
              ||| |||     |||  |  |        |     |  |  ||  ||    |   ||
Db        416 CTCAACCAATTCCTCTTCTCACCCATGGCCATACGGTTTCTGGAGAGATTCGCACGCCTG 475

Qy        489 CTGATAGACATGCTCTTATTGTTCCTCCTTCACTTGGTGGTCATGGCAATAGAGTTCATC 548
              |    |  ||  |  |  |  |  |  | ||| |  || ||  |  |  |
Db        476 ATACACAATCTGTGCGAACTACATCAGGTCCTTTGGGTCCTTCTGACAGGAATGCTATTT 535

Qy        549 CTGTTTCTCTTTCTGACCCGACCGTGGCTGCACATCGAAGGCTGATGGTACCTCAGAAAG 608
              |   |  |   |  ||| ||         |||  |  |  |||  |  |||  ||    ||||
Db        536 CATCTCCATATATTGATCCACGGCAACCTGTCCCTGTAAGAATCGTGGACCCGTCAAAAG 595

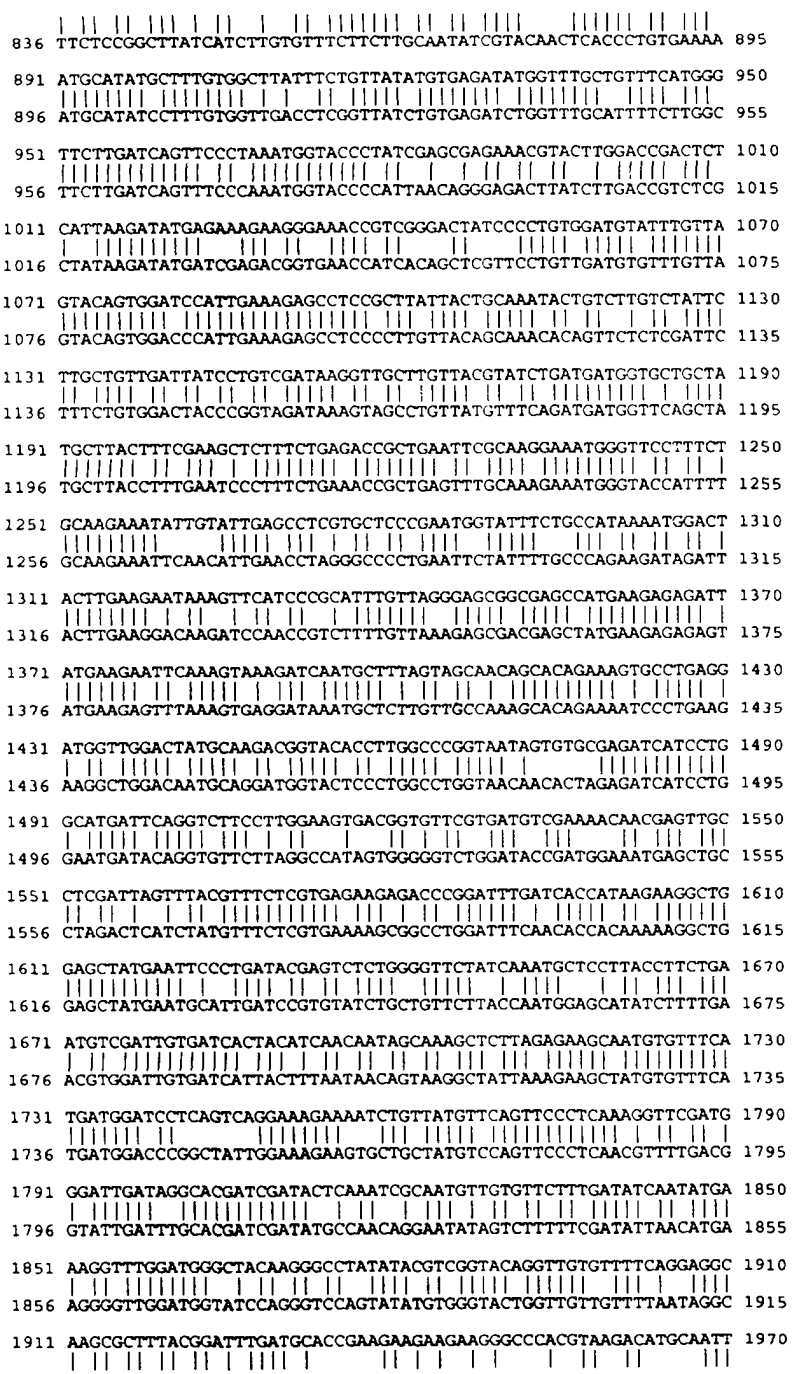
Qy        609 ATCTTGCGGTTTATGGTTATGGAAGTGTCGCTTGGAAAGATCGGATGGAGGAATGGAAGA 668
              |  |      ||||||   |||  | |||  |  ||||||||  |  |  || |  ||||||
Db        596 ACTTGAACTCTTATGGGCTTGGTAATGTTGACTGGAAAGAAAGAGTTGAAGGCTGGAAGC 655

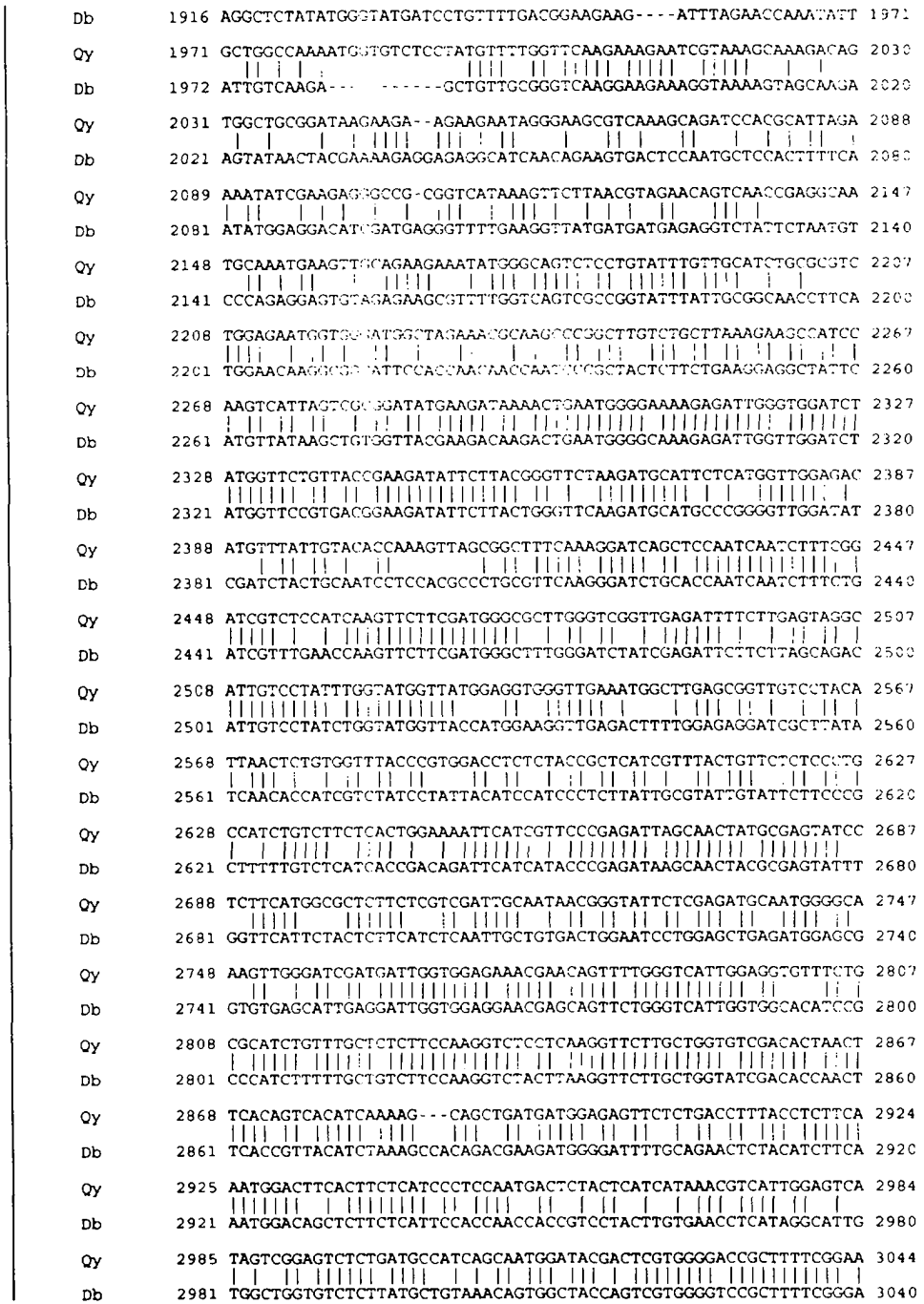
Qy        669 GAAAGCAGAATGAGAAACT-----------------TCAGGTTGTTAGGCATGAAGGAG 710
                |||||| |  | ||  |                    |   ||  || | |  |||||
Db        656 TGAAGCAGGAGAAAAATATGTTACAGATGACTGGTAAATACCATGAAGGGAAAGGAGGAG 715

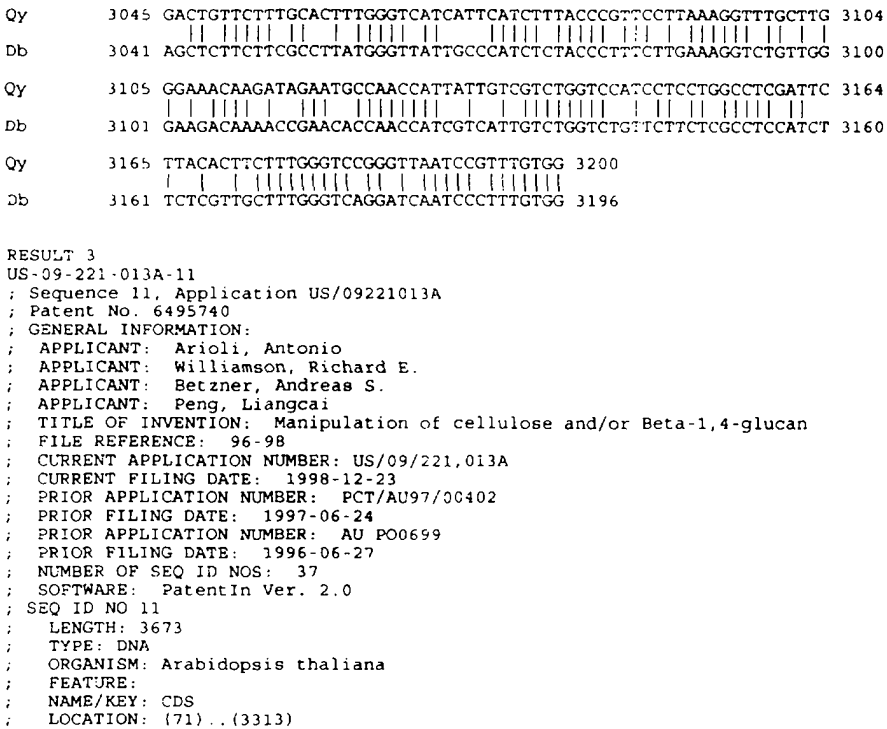
Qy        711 ATCCTGATTTTGAAGATGGTGATGATGCTGATTTTCCAATGATGGATGAGGGAAGGCAGC 770
              |  |||       | |    |||  |  ||  | | ||||   |||||     |  | |  |
Db        716 AAATTGAAGGGACTGGTTCCAATGGCGAAGAACTCCAAATGGCTGATGATACACGTCTTC 775

Qy        771 CATTGTCTATGAAGATACCAATCAAATCGAGCAAGATAAATCCTTACCGGATGTTAATTG 830
              |  ||  |     | | ||  |||   |||      |||  ||||| ||| |  | |||
Db        776 CTATGAGTCGTGTGGTGCCTATCCCATCTTCTCGCCTAACCCCTTATCGGGTTGTGATTA 835

Qy        831 TGCTACGTCTTGTGATTCTTGGTCTCTTCTTTCACTACCGTATTCTTCACCCCGTCAAAG 890
```

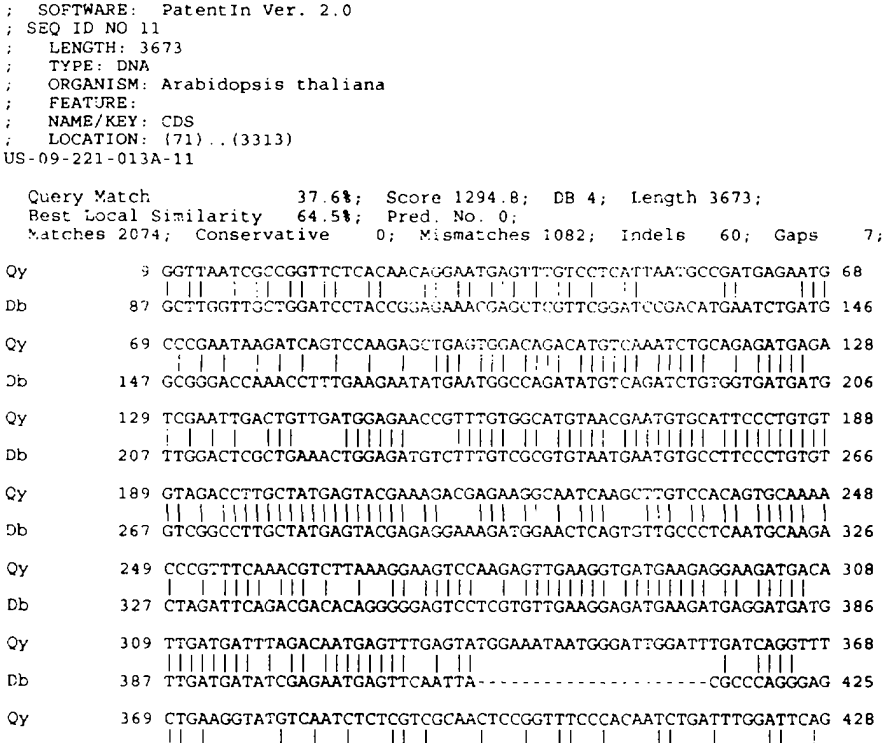

836 TTCTCCGGCTTATCATCTTGTGTTTCTTCTTGCAATATCGTACAACTCACCCTGTGAAAA 895
891 ATGCATATGCTTTGTGGCTTATTTCTGTTATATGTGAGATATGGTTTGCTGTTTCATGGG 950
896 ATGCATATCCTTTGTGGTTGACCTCGGTTATCTGTGAGATCTGGTTTGCATTTTCTTGGC 955
951 TTCTTGATCAGTTCCCTAAATGGTACCCTATCGAGCGAGAAACGTACTTGGACCGACTCT 1010
956 TTCTTGATCAGTTTCCCAAATGGTACCCCATTAACAGGGAGACTTATCTTGACCGTCTCG 1015
1011 CATTAAGATATGAGAAAGAAGGGAAACCGTCGGGACTATCCCCTGTGGATGTATTTGTTA 1070
1016 CTATAAGATATGATCGAGACGGTGAACCATCACAGCTCGTTCCTGTTGATGTGTTTGTTA 1075
1071 GTACAGTGGATCCATTGAAAGAGCCTCCGCTTATTACTGCAAATACTGTCTTGTCTATTC 1130
1076 GTACAGTGGACCCATTGAAAGAGCCTCCCCTTGTTACAGCAAACACAGTTCTCTCGATTC 1135
1131 TTGCTGTTGATTATCCTGTCGATAAGGTTGCTTGTTACGTATCTGATGATGGTGCTGCTA 1190
1136 TTTCTGTGGACTACCCGGTAGATAAAGTAGCCTGTTATGTTTCAGATGATGGTTCAGCTA 1195
1191 TGCTTACTTTCGAAGCTCTTTCTGAGACCGCTGAATTCGCAAGGAAATGGGTTCCTTTCT 1250
1196 TGCTTACCTTTGAATCCCTTTCTGAAACCGCTGAGTTTGCAAAGAAATGGGTACCATTTT 1255
1251 GCAAGAAATATTGTATTGAGCCTCGTGCTCCCGAATGGTATTTCTGCCATAAAATGGACT 1310
1256 GCAAGAAATTCAACATTGAACCTAGGGCCCCTGAATTCTATTTTGCCCAGAAGATAGATT 1315
1311 ACTTGAAGAATAAAGTTCATCCCGCATTTGTTAGGGAGCGGCGAGCCATGAAGAGAGATT 1370
1316 ACTTGAAGGACAAGATCCAACCGTCTTTTGTTAAAGAGCGACGAGCTATGAAGAGAGAGT 1375
1371 ATGAAGAATTCAAAGTAAAGATCAATGCTTTAGTAGCAACAGCACAGAAAGTGCCTGAGG 1430
1376 ATGAAGAGTTTAAAGTGAGGATAAATGCTCTTGTTGCCAAAGCACAGAAAATCCCTGAAG 1435
1431 ATGGTTGGACTATGCAAGACGGTACACCTTGGCCCGGTAATAGTGTGCGAGATCATCCTG 1490
1436 AAGGCTGGACAATGCAGGATGGTACTCCCTGGCCTGGTAACAACACTAGAGATCATCCTG 1495
1491 GCATGATTCAGGTCTTCCTTGGAAGTGACGGTGTTCGTGATGTCGAAAACAACGAGTTGC 1550
1496 GAATGATACAGGTGTTCTTAGGCCATAGTGGGGGTCTGGATACCGATGGAAATGAGCTGC 1555
1551 CTCGATTAGTTTACGTTTCTCGTGAGAAGAGACCCGGATTTGATCACCATAAGAAGGCTG 1610
1556 CTAGACTCATCTATGTTTCTCGTGAAAAGCGGCCTGGATTTCAACACCACAAAAAGGCTG 1615
1611 GAGCTATGAATTCCCTGATACGAGTCTCTGGGGTTCTATCAAATGCTCCTTACCTTCTGA 1670
1616 GAGCTATGAATGCATTGATCCGTGTATCTGCTGTTCTTACCAATGGAGCATATCTTTTGA 1675
1671 ATGTCGATTGTGATCACTACATCAACAATAGCAAAGCTCTTAGAGAAGCAATGTGTTTCA 1730
1676 ACGTGGATTGTGATCATTACTTTAATAACAGTAAGGCTATTAAAGAAGCTATGTGTTTCA 1735
1731 TGATGGATCCTCAGTCAGGAAAGAAAATCTGTTATGTTCAGTTCCCTCAAAGGTTCGATG 1790
1736 TGATGGACCCGGCTATTGGAAAGAAGTGCTGCTATGTCCAGTTCCCTCAACGTTTTGACG 1795
1791 GGATTGATAGGCACGATCGATACTCAAATCGCAATGTTGTGTTCTTTGATATCAATATGA 1850
1796 GTATTGATTTGCACGATCGATATGCCAACAGGAATATAGTCTTTTTCGATATTAACATGA 1855
1851 AAGGTTTGGATGGGCTACAAGGGCCTATATACGTCGGTACAGGTTGTGTTTTCAGGAGGC 1910
1856 AGGGGTTGGATGGTATCCAGGGTCCAGTATATGTGGGTACTGGTTGTTGTTTTAATAGGC 1915
1911 AAGCGCTTTACGGATTTGATGCACCGAAGAAGAAGAAGGGCCCACGTAAGACATGCAATT 1970


Db 1916 AGGCTCTATATGGGTATGATCCTGTTTTGACGGAAGAAG----ATTTAGAACCAAATATT 1971
Qy 1971 GCTGGCCAAAATGGTGTCTCCTATGTTTTGGTTCAAGAAAGAATCGTAAAGCAAAGACAG 2030
Db 1972 ATTGTCAAGA--- -------GCTGTTGCGGGTCAAGGAAGAAAGGTAAAAGTAGCAAGA 2020
Qy 2031 TGGCTGCGGATAAGAAGA--AGAAGAATAGGGAAGCGTCAAAGCAGATCCACGCATTAGA 2088
Db 2021 AGTATAACTACGAAAAGAGGAGAGGCATCAACAGAAGTGACTCCAATGCTCCACTTTTCA 2080
Qy 2089 AAATATCGAAGAGGGCCG-CGGTCATAAAGTTCTTAACGTAGAACAGTCAACCGAGGCAA 2147
Db 2081 ATATGGAGGACATCGATGAGGGTTTTGAAGGTTATGATGATGAGAGGTCTATTCTAATGT 2140
Qy 2148 TGCAAATGAAGTTGCAGAAGAAATATGGGCAGTCTCCTGTATTTGTTGCATCTGCGCGTC 2207
Db 2141 CCCAGAGGAGTGTAGAGAAGCGTTTTGGTCAGTCGCCGGTATTTATTGCGGCAACCTTCA 2200
Qy 2208 TGGAGAATGGTGG[illegible]ATGGCTAGAAACGCAAGCCCGGCTTGTCTGCTTAAAGAAGCCATCC 2267
Db 2201 TGGAACAAGGCGG[illegible]ATTCCACCAACAACCAACTCCGCTACTCTTCTGAAGGAGGCTATTC 2260
Qy 2268 AAGTCATTAGTCGCGGATATGAAGATAAAACTGAATGGGGAAAAGAGATTGGGTGGATCT 2327
Db 2261 ATGTTATAAGCTGTGGTTACGAAGACAAGACTGAATGGGGCAAAGAGATTGGTTGGATCT 2320
Qy 2328 ATGGTTCTGTTACCGAAGATATTCTTACGGGTTCTAAGATGCATTCTCATGGTTGGAGAC 2387
Db 2321 ATGGTTCCGTGACGGAAGATATTCTTACTGGGTTCAAGATGCATGCCCGGGGTTGGATAT 2380
Qy 2388 ATGTTTATTGTACACCAAAGTTAGCGGCTTTCAAAGGATCAGCTCCAATCAATCTTTCGG 2447
Db 2381 CGATCTACTGCAATCCTCCACGCCCTGCGTTCAAGGGATCTGCACCAATCAATCTTTCTG 2440
Qy 2448 ATCGTCTCCATCAAGTTCTTCGATGGGCGCTTGGGTCGGTTGAGATTTTCTTGAGTAGGC 2507
Db 2441 ATCGTTTGAACCAAGTTCTTCGATGGGCTTTGGGATCTATCGAGATTCTTCTTAGCAGAC 2500
Qy 2508 ATTGTCCTATTTGGTATGGTTATGGAGGTGGGTTGAAATGGCTTGAGCGGTTGTCCTACA 2567
Db 2501 ATTGTCCTATCTGGTATGGTTACCATGGAAGGTTGAGACTTTTGGAGAGGATCGCTTATA 2560
Qy 2568 TTAACTCTGTGGTTTACCCGTGGACCTCTCTACCGCTCATCGTTTACTGTTCTCTCCCTG 2627
Db 2561 TCAACACCATCGTCTATCCTATTACATCCATCCCTCTTATTGCGTATTGTATTCTTCCCG 2620
Qy 2628 CCATCTGTCTTCTCACTGGAAAATTCATCGTTCCCGAGATTAGCAACTATGCGAGTATCC 2687
Db 2621 CTTTTTGTCTCATCACCGACAGATTCATCATACCCGAGATAAGCAACTACGCGAGTATTT 2680
Qy 2688 TCTTCATGGCGCTCTTCTCGTCGATTGCAATAACGGGTATTCTCGAGATGCAATGGGGCA 2747
Db 2681 GGTTCATTCTACTCTTCATCTCAATTGCTGTGACTGGAATCCTGGAGCTGAGATGGAGCG 2740
Qy 2748 AAGTTGGGATCGATGATTGGTGGAGAAACGAACAGTTTTGGGTCATTGGAGGTGTTTCTG 2807
Db 2741 GTGTGAGCATTGAGGATTGGTGGAGGAACGAGCAGTTCTGGGTCATTGGTGGCACATCCG 2800
Qy 2808 CGCATCTGTTTGCTCTCTTCCAAGGTCTCCTCAAGGTTCTTGCTGGTGTCGACACTAACT 2867
Db 2801 CCCATCTTTTTGCTGTCTTCCAAGGTCTACTTAAGGTTCTTGCTGGTATCGACACCAACT 2860
Qy 2868 TCACAGTCACATCAAAAG---CAGCTGATGATGGAGAGTTCTCTGACCTTTACCTCTTCA 2924
Db 2861 TCACCGTTACATCTAAAGCCACAGACGAAGATGGGGATTTTGCAGAACTCTACATCTTCA 2920
Qy 2925 AATGGACTTCACTTCTCATCCCTCCAATGACTCTACTCATCATAAACGTCATTGGAGTCA 2984
Db 2921 AATGGACAGCTCTTCTCATTCCACCAACCACCGTCCTACTTGTGAACCTCATAGGCATTG 2980
Qy 2985 TAGTCGGAGTCTCTGATGCCATCAGCAATGGATACGACTCGTGGGGACCGCTTTTCGGAA 3044
Db 2981 TGGCTGGTGTCTCTTATGCTGTAAACAGTGGCTACCAGTCGTGGGGTCCGCTTTTCGGGA 3040



RESULT 3
US-09-221-013A-11
; Sequence 11, Application US/09221013A
; Patent No. 6495740
; GENERAL INFORMATION:
; APPLICANT: Arioli, Antonio
; APPLICANT: Williamson, Richard E.
; APPLICANT: Betzner, Andreas S.
; APPLICANT: Peng, Liangcai
; TITLE OF INVENTION: Manipulation of cellulose and/or Beta-1,4-glucan
; FILE REFERENCE: 96-98
; CURRENT APPLICATION NUMBER: US/09/221,013A
; CURRENT FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: PCT/AU97/00402
; PRIOR FILING DATE: 1997-06-24
; PRIOR APPLICATION NUMBER: AU PO0699
; PRIOR FILING DATE: 1996-06-27
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 3673
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (71)..(3313)
US-09-221-013A-11

Query Match 37.6%; Score 1294.8; DB 4; Length 3673;
Best Local Similarity 64.5%; Pred. No. 0;
Matches 2074; Conservative 0; Mismatches 1082; Indels 60; Gaps 7;



```
Qy      1491 GCATGATTCAGGTCTTCCTTGGAAGTGACGGTGTTCGTGATGTCGAAAACAACGAGTTGC 1550
Db      1566 GAATGATACAGGTGTTCTTAGGCCATAGTGGGGGTCTGGATACCGATGGAAATGAGCTGC 1625

Qy      1551 CTCGATTAGTTTACGTTTCTCGTGAGAAGAGACCCGGATTTGATCACCATAAGAAGGCTG 1610
Db      1626 CTAGACTCATCTATGTTTCTCGTGAAAAGCGGCCTGGATTTCAACACCACAAAAAGGCTG 1685

Qy      1611 GAGCTATGAATTCCCTGATACGAGTCTCTGGGGTTCTATCAAATGCTCCTTACCTTCTGA 1670
Db      1686 GAGCTATGAATGCATTGATCCGTGTATCTGTTGTTCTTACCAATGGAGCATATCTTTTGA 1745

Qy      1671 ATGTCGATTGTGATCACTACATCAACAATAGCAAAGCTCTTAGAGAAGCAATGTGTTTCA 1730
Db      1746 ACGTGGATTGTGATCATTACTTTAATAACAGTAAGGCTATTAAAGAAGCTATGTGTTTCA 1805

Qy      1731 TGATGGATCCTCAGTCAGGAAAGAAAATCTGTTATGTTCAGTTCCCTCAAAGGTTCGATG 1790
Db      1806 TGATGGACCCGGCTATTGGAAAGAAGTGCTGCTATGTCCAGTTCCCTCAACGTTTTGACG 1865

Qy      1791 GGATTGATAGGCACGATCGATACTCAAATCGCAATGTTGTGTTCTTTGATATCAATATGA 1850
Db      1866 GTATTGATTTGCACGATCGATATGCCAACAGGAATATAGTCTTTTTCGATATTAACATGA 1925

Qy      1851 AAGGTTTGGATGGGCTACAAGGGCCTATATACGTCGGTACAGGTTGTGTTTTCAGGAGGC 1910
Db      1926 AGGGGTTGGATGGTATCCAGGGTCCAGTATATGTGGGTACTGGTTGTTGTTTTAATAGGC 1985

Qy      1911 AAGCGCTTTACGGATTTGATGCACCGAAGAAGAAGAAGGGCCCACGTAAGACATGCAATT 1970
Db      1986 AGGCTCTATATGGGTATGATCCTGTTTTGACGGAAGAAG----ATTTAGAACCAAATATT 2041

Qy      1971 GCTGGCCAAAATGGTGTCTCCTATGTTTTGGTTCAAGAAAGAATCGTAAAGCAAAGACAG 2030
Db      2042 ATTGTCAAGA-----------GCTGTTGCGGGTCAAGGAAGAAAGGTAAAAGTAGCAAGA 2090

Qy      2031 TGGCTGCGGATAAGAAGA--AGAAGAATAGGGAAGCGTCAAAGCAGATCCACGCATTAGA 2088
Db      2091 AGTATAACTACGAAAAGAGGAGAGGCATCAACAGAAGTGACTCCAATGCTCCACTTTTCA 2150

Qy      2089 AAATATCGAAGAGGGCCG-CGGTCATAAAGTTCTTAACGTAGAACAGTCAACCGAGGCAA 2147
Db      2151 ATATGGAGGACATCGATGAGGGTTTTGAAGGTTATGATGATGAGAGGTCTATTCTAATGT 2210

Qy      2148 TGCAAATGAAGTTGCAGAAGAAATATGGGCAGTCTCCTGTATTTGTTGCATCTGCGCGTC 2207
Db      2211 CCCAGAGGAGTGTAGAGAAGCGTTTTGGTCAGTCGCCGGTATTTATTGCGGCAACCTTCA 2270

Qy      2208 TGGAGAATGGTGGGATGGCTAGAAACGCAAGCCCGGCTTGTCTGCTTAAAGAAGCCATCC 2267
Db      2271 TGGAACAAGGCGGCATTCCACCAACAACCAATCCCGCTACTCTTCTGAAGGAGGCTATTC 2330

Qy      2268 AAGTCATTAGTCGCGGATATGAAGATAAAACTGAATGGGGAAAAGAGATTGGGTGGATCT 2327
Db      2331 ATGTTATAAGCTGTGGTTACGAAGACAAGACTGAATGGGGCAAAGAGATTGGTTGGATCT 2390

Qy      2328 ATGGTTCTGTTACCGAAGATATTCTTACGGGTTCTAAGATGCATTCTCATGGTTGGAGAC 2387
Db      2391 ATGGTTCCGTGACGGAAGATATTCTTACTGGGTTCAAGATGCATGCCCGGGGTTGGATAT 2450

Qy      2388 ATGTTTATTGTACACCAAAGTTAGCGGCTTTCAAAGGATCAGCTCCAATCAATCTTTCGG 2447
Db      2451 CGATCTACTGCAATCCTCCACGCCCTGCGTTCAAGGGATCTGCACCAATCAATCTTTCTG 2510

Qy      2448 ATCGTCTCCATCAAGTTCTTCGATGGGCGCTTGGGTCGGTTGAGATTTTCTTGAGTAGGC 2507
Db      2511 ATCGTTTGAACCAAGTTCTTCGATGGGCTTTGGGATCTATCGAGATTCTTCTTAGCAGAC 2570

Qy      2508 ATTGTCCTATTTGGTATGGTTATGGAGGTGGGTTGAAATGGCTTGAGCGGTTGTCCTACA 2567
Db      2571 ATTGTCCTATCTGGTATGGTTACCATGGAAGGTTGAGACTTTTGGAGAGGATCGCTTATA 2630

Qy      2568 TTAACTCTGTGGTTTACCCGTGGACCTCTCTACCGCTCATCGTTTACTGTTCTCTCCCTG 2627
```

```
Db      2631 TCAACACCATCGTCTATCCTATTACATCCATCCCTCTTATTGCGTATTGTATTCTTCCCG 2690

Qy      2628 CCATCTGTCTTCTCACTGGAAAATTCATCGTTCCCGAGATTAGCAACTATGCGAGTATCC 2687
Db      2691 CTTTTTGTCTCATCACCGACAGATTCATCATACCCGAGATAAGCAACTACGCGAGTATTT 2750

Qy      2688 TCTTCATGGCGCTCTTCTCGTCGATTGCAATAACGGGTATTCTCGAGATGCAATGGGGCA 2747
Db      2751 GGTTCATTCTACTCTTCATCTCAATTGCTGTGACTGGAATCCTGGAGCTGAGATGGAGCG 2810

Qy      2748 AAGTTGGGATCGATGATTGGTGGAGAAACGAACAGTTTTGGGTCATTGGAGGTGTTTCTG 2807
Db      2811 GTGTGAGCATTGAGGATTGGTGGAGGAACGAGCAGTTCTGGGTCATTGGTGGCACATCCG 2870

Qy      2808 CGCATCTGTTTGCTCTCTTCCAAGGTCTCCTCAAGGTTCTTGCTGGTGTCGACACTAACT 2867
Db      2871 CCCATCTTTTTGCTGTCTTCCAAGGTCTACTTAAGGTTCTTGCTGGTATCGACACCAACT 2930

Qy      2868 TCACAGTCACATCAAAAG---CAGCTGATGATGGAGAGTTCTCTGACCTTTACCTCTTCA 2924
Db      2931 TCACCGTTACATCTAAAGCCACAGACGAAGATGGGATTTTGCAGAACTCTACATCTTCA 2990

Qy      2925 AATGGACTTCACTTCTCATCCCTCCAATGACTCTACTCATCATAAACGTCATTGGAGTCA 2984
Db      2991 AATGGACAGCTCTTCTCATTCCACCAACCACCGTCCTACTTGTGAACCTCATAGGCATTG 3050

Qy      2985 TAGTCGGAGTCTCTGATGCCATCAGCAATGGATACGACTCGTGGGGACCGCTTTTCGGAA 3044
Db      3051 TGGCTGGTGTCTCTTATGCTGTAAACAGTGGCTACCAGTCGTGGGGTCCGCTTTTCGGGA 3110

Qy      3045 GACTGTTCTTTGCACTTTGGGTCATCATTCATCTTTACCCGTTCCTTAAAGGTTTGCTTG 3104
Db      3111 AGCTCTTCTTCGCCTTATGGGTTATTGCCCATCTCTACCCTTTCTTGAAAGGTCTGTTGG 3170

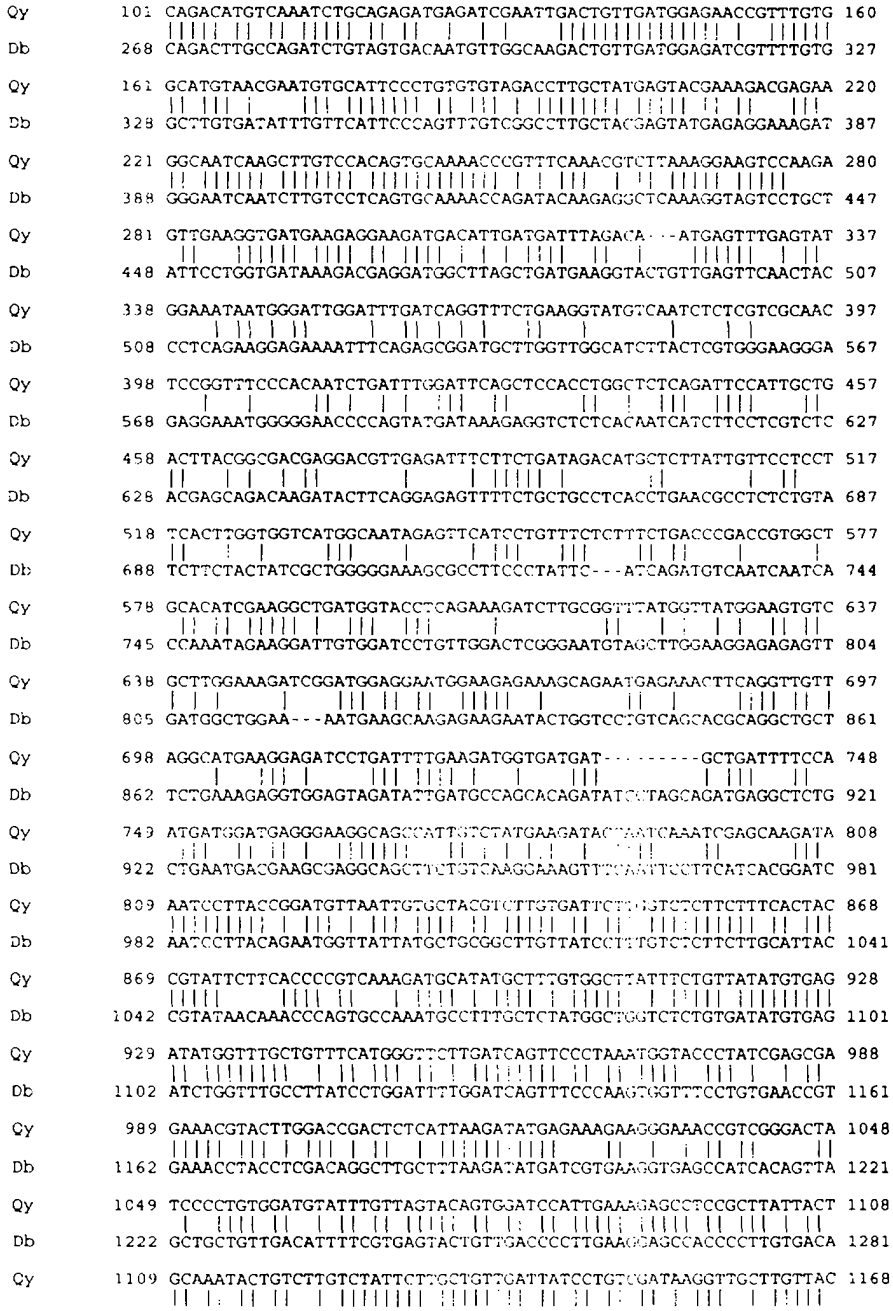
Qy      3105 GGAAACAAGATAGAATGCCAACCATTATTGTCGTCTGGTCCATCCTCCTGGCCTCGATTC 3164
Db      3171 GAAGACAAAACCGAACACCAACCATCGTCATTGTCTGGTCTGTTCTTCTCGCCTCCATCT 3230

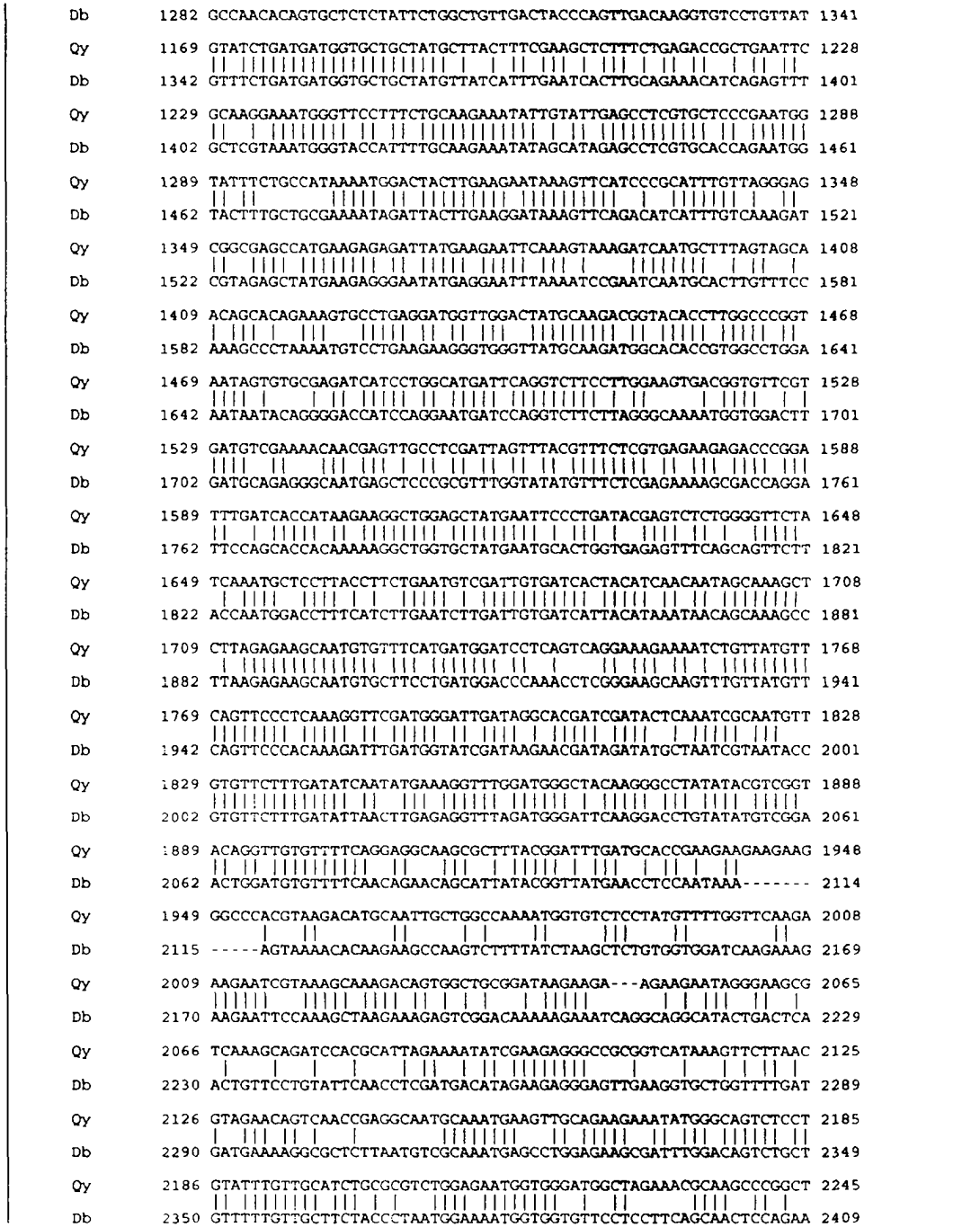
Qy      3165 TTACACTTCTTTGGGTCCGGGTTAATCCGTTTGTGG 3200
Db      3231 TCTCGTTGCTTTGGGTCAGGATCAATCCCTTTGTGG 3266
```

```
RESULT 4
US-09-221-013A-9
; Sequence 9, Application US/09221013A
; Patent No. 6495740
; GENERAL INFORMATION:
;  APPLICANT:  Arioli, Antonio
;  APPLICANT:  Williamson, Richard E.
;  APPLICANT:  Betzner, Andreas S.
;  APPLICANT:  Peng, Liangcai
;  TITLE OF INVENTION:  Manipulation of cellulose and/or Beta-1,4-glucan
;  FILE REFERENCE:  96-98
;  CURRENT APPLICATION NUMBER:  US/09/221,013A
;  CURRENT FILING DATE:  1998-12-23
;  PRIOR APPLICATION NUMBER:  PCT/AU97/00402
;  PRIOR FILING DATE:  1997-06-24
;  PRIOR APPLICATION NUMBER:  AU PO0699
;  PRIOR FILING DATE:  1996 06-27
;  NUMBER OF SEQ ID NOS:  37
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 9
;  LENGTH: 3614
;  TYPE: DNA
;  ORGANISM: Arabidopsis thaliana
;  FEATURE:
;  NAME/KEY: CDS
;  LOCATION: (217)..(3411)
US-09-221-013A-9

  Query Match            36.4%;  Score 1252.6;  DB 4;  Length 3614;
```

Best Local Similarity 64.3%; Pred. No. 0;
Matches 2003; Conservative 0; Mismatches 1074; Indels 36; Gaps 7;





```
Qy      2246 TGTCTGCTTAAAGAAGCCATCCAAGTCATTAGTCGCGGATATGAAGATAAAACTGAATGG 2305
Db      2410 AACTTTCTCAAAGAGGCTATCCATGTCATTAGTTGTGGTTATGAGGATAAGTCAGATTGG 2469

Qy      2306 GGAAAAGAGATTGGGTGGATCTATGGTTCTGTTACCGAAGATATTCTTACGGGTTCTAAG 2365
Db      2470 GGAATGGAGATTGGATGGATCTATGGTTCTGTGACAGAAGATATTCTGACTGGGTTCAAA 2529

Qy      2366 ATGCATTCTCATGGTTGGAGACATGTTTATTGTACACCAAAGTTAGCGGCTTTCAAAGGA 2425
Db      2530 ATGCATGCCCGTGGATGGCGATCCATTTACTGCATGCCTAAGCTTCCAGCTTTCAAGGGT 2589

Qy      2426 TCAGCTCCAATCAATCTTTCGGATCGTCTCCATCAAGTTCTTCGATGGGCGCTTGGGTCG 2485
Db      2590 TCTGCTCCTATCAATCTTTCAGATCGTCTGAACCAAGTGCTGAGGTGGGCTTTAGGTTCA 2649

Qy      2486 GTTGAGATTTTCTTGAGTAGGCATTGTCCTATTTGGTATGGTTATGGAGGTGGGTTGAAA 2545
Db      2650 GTTGAGATTCTCTTCAGTCGGCATTGTCCTATATGGTATGGTTACAATGGGAGGCTAAAA 2709

Qy      2546 TGGCTTGAGCGGTTGTCCTACATTAACTCTGTGGTTTACCCGTGGACCTCTCTACCGCTC 2605
Db      2710 TTTCTTGAGAGGTTTGCGTATGTGAACACCACCATCTACCCTATCACCTCCATTCCTCTT 2769

Qy      2606 ATCGTTTACTGTTCTCTCCCTGCCATCTGTCTTCTCACTGGAAAATTCATCGTTCCCGAG 2665
Db      2770 CTCATGTATTGTACATTGCTAGCCGTTTGTCTCTTCACCAACCAGTTTATTATTCCTCAG 2829

Qy      2666 ATTAGCAACTATGCGAGTATCCTCTTCATGGCGCTCTTCTCGTCGATTGCAATAACGGGT 2725
Db      2830 ATTAGTAACATTGCAAGTATATGGTTTCTGTCTCTCTTTCTCTCCATTTTCGCCACGGGT 2889

Qy      2726 ATTCTCGAGATGCAATGGGGCAAAGTTGGGATCGATGATTGGTGGAGAAACGAACAGTTT 2785
Db      2890 ATACTAGAAATGAGGTGGAGTGGCGTAGGCATAGACGAATGGTGGAGAAACGAGCAGTTT 2949

Qy      2786 TGGGTCATTGGAGGTGTTTCTGCGCATCTGTTTGCTCTCTTCCAAGGTCTCCTCAAGGTT 2845
Db      2950 TGGGTCATTGGTGGAGTATCCGCTCATTTATTCGCTGTGTTTCAAGGTATCCTCAAAGTC 3009

Qy      2846 CTTGCTGGTGTCGACACTAACTTCACAGTCACATCAAAAG---CAGCTGATGATGGAGAG 2902
Db      3010 CTTGCCGGTATTGACACAAACTTCACAGTTACCTCAAAAGCTTCAGATGAAGACGGAGAC 3069

Qy      2903 TTCTCTGACCTTTACCTCTTCAAATGGACTTCACTTCTCATCCCTCCAATGACTCTACTC 2962
Db      3070 TTTGCTGAGCTCTACTTGTTCAAATGGACAACACTTCTGATTCCGCCAACGACGCTGCTC 3129

Qy      2963 ATCATAAACGTCATTGGAGTCATAGTCGGAGTCTCTGATGCCATCAGCAATGGATACGAC 3022
Db      3130 ATTGTAAACTTAGTGGGAGTTGTTGCAGGAGTCTCTTATGCTATCAACAGTGGATACCAA 3189

Qy      3023 TCGTGGGGACCGCTTTTCGGAAGACTGTTCTTTGCACTTTGGGTCATCATTCATCTTTAC 3082
Db      3190 TCATGGGGACCACTCTTTGGTAAGTTGTTCTTTGCCTTCTGGGTGATTGTTCACTTGTAC 3249

Qy      3083 CCGTTCCTTAAAGGTTTGCTTGGGAAACAAGATAGAATGCCAACCATTATTGTCGTCTGG 3142
Db      3250 CCTTTCCTCAAGGGTTTGATGGGTCGACAGAACCGGACTCCTACCATTGTTGTGGTCTGG 3309

Qy      3143 TCCATCCTCCTGGCCTCGATTCTTACACTTCTTTGGGTCCGGGTTAATCCGTT 3195
Db      3310 TCTGTTCTCTTGGCTTCTATCTTCTCGTTGTTGTGGGTTAGGATTGATCCCTT 3362
```

```
ESULT 5
IS-08-960-048-1
  Sequence 1, Application US/08960048C
  Patent No. 6271443
  GENERAL INFORMATION:
   APPLICANT:  Stalker, D. et al.
   TITLE OF INVENTION:  Plant Cellulose Synthase and Promoter
```

```
;  TITLE OF INVENTION:  Sequences
;  FILE REFERENCE:  15621/01/US
;  CURRENT APPLICATION NUMBER:  US/08/960,048C
;  CURRENT FILING DATE:  1997-10-29
;  PRIOR APPLICATION NUMBER:  60/029,987
;  PRIOR FILING DATE:  1996-10-29
;  NUMBER OF SEQ ID NOS:  12
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;  LENGTH:  3328
;  TYPE:  DNA
;  ORGANISM:  Artificial Sequence
;  FEATURE:
;  OTHER INFORMATION:  Synthetic Oligonucleotide
US-08-960-048-1

  Query Match             [illegible]%;  Score 1145;  DB 3;  Length 3328;
  Best Local Similarity   [illegible]%;  Pred. No. 0.
  Matches 1721;  Conservative    0;  Mismatches  845;  Indels   39;  Gaps    3;

Qy       616 GGTTTATG[illegible]TAG[illegible]AAGTGT[illegible]GGAAAGATCGGATGGAGGAATGGAAGAGAAAGCA 675
Db       487 GGCTGAAGA[illegible]AATGGGAATTCGATTTGGAAGAACAGGGTGGAAAGTTGGAAAGAAAAGAA 546

Qy       676 GAATGAGAAACTTCAGGTTG------TTAGGCATGAAGGAGATCCTGATTTTGAAGATGG 729
Db       547 GAACAAGAAGAAGAAGCCTGCAACAACTAAGGTTGAAAGAGAGGCTGAAATCCCACCTGA 606

Qy       730 TGATGATGCTGATTTTCCAATGATGGATGAGGGAAGGCAGCCATTGTCTATGAAGATACC 789
Db       607 GCAACAAATGGAAGATAAACCGGCACCGGATGCTTCCCAGCCCCTCTCGACTATAATTCC 666

Qy       790 AATCAAATCGAGCAAGATAAATCCTTACCGGATGTTAATTGTGCTACGTCTTGTGATTCT 849
Db       667 AATCCCGAAAAGCAGACTTGCACCATACCGAACCGTGATCATTATGCGATTGATCATTCT 726

Qy       850 TGGTCTCTTCTTTCACTACCGTATTCTTCACCCCGTCAAAGATGCATATGCTTTGTGGCT 909
Db       727 TGGTCTTTTCTTCCATTATCGAGTAACAAACCCCGTTGACAGTGCTTTTGGACTGTGGCT 786

Qy       910 TATTTCTGTTATATGTGAGATATGGTTTGCTGTTTCATGGGTTCTTGATCAGTTCCCTAA 969
Db       787 CACTTCAGTCATATGTGAAATCTGGTTTGCATTTTCCTGGGTGTTGGATCAGTTCCCTAA 846

Qy       970 ATGGTACCCTATCGAGCGAGAAACGTACTTGGACCGACTCTCATTAAGATATGAGAAAGA 1029
Db       847 GTGGTATCCTGTTAACAGGGAAACATACATTGACAGACTATCTGCAAGATATGAAAGAGA 906

Qy      1030 AGGGAAACCGTCGGGACTATCCCCTGTGGATGTATTTGTTAGTACAGTGGATCCATTGAA 1089
Db       907 AGGTGAACCTGATGAACTTGCTGCAGTTGACTTCTTCGTGAGTACAGTGGATCCATTGAA 966

Qy      1090 AGAGCCTCCGCTTATTACTGCAAATACTGTCTTGTCTATTCTTGCTGTTGATTATCCTGT 1149
Db       967 AGAGCCTCCATTGATTACTGCCAATACTGTGCTTTCCATCCTTGCCTTGGACTACCCGGT 1026

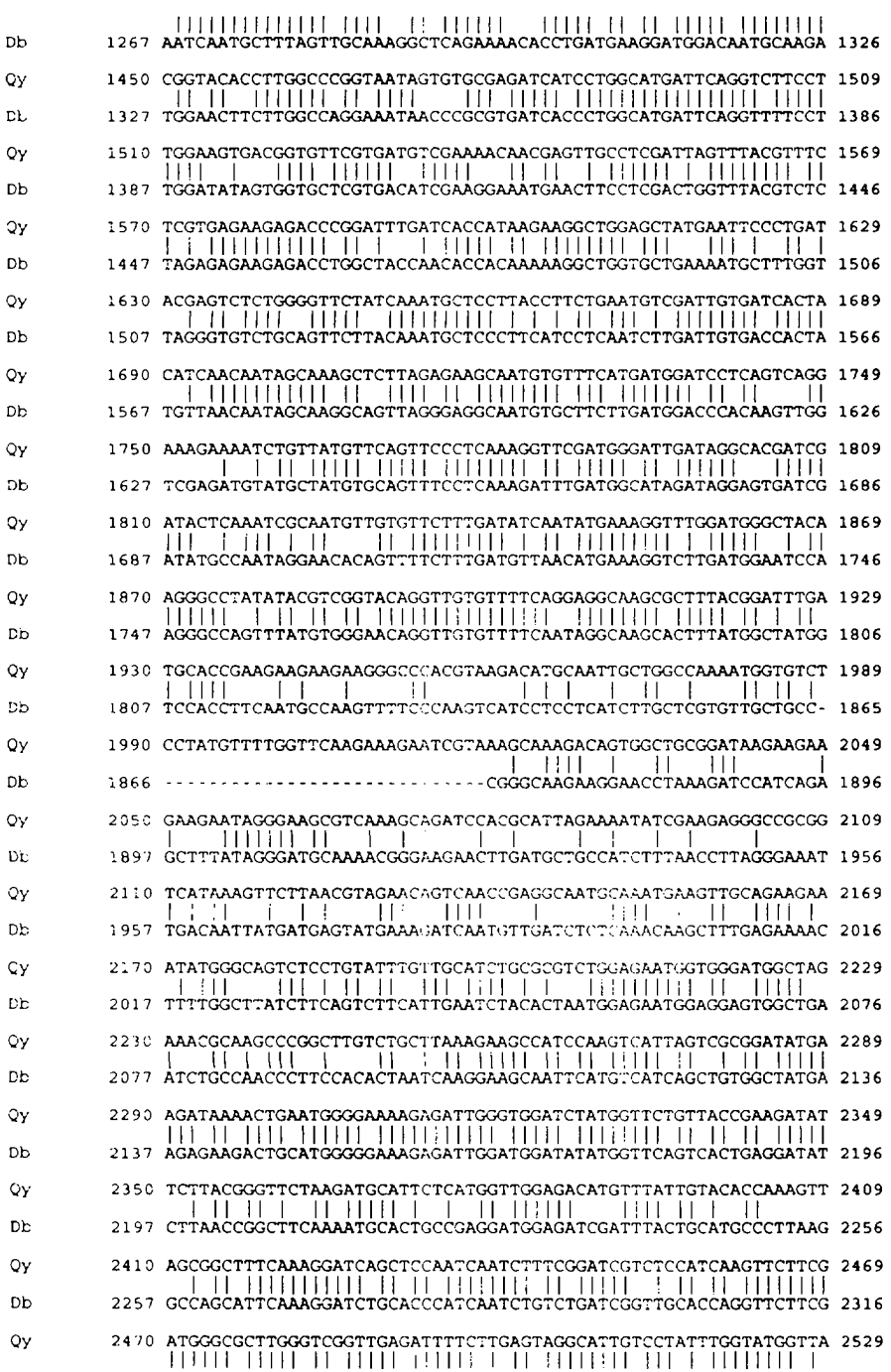
Qy      1150 CGATAAGGTTGCTTGTTACGTATCTGATGATGGTGCTGCTATGCTTACTTTCGAAGCTCT 1209
Db      1027 GGATAAGGTCTCTTGTTATATATCTGATGATGGTGCGGCCATGCTGACATTTGAATCTCT 1086

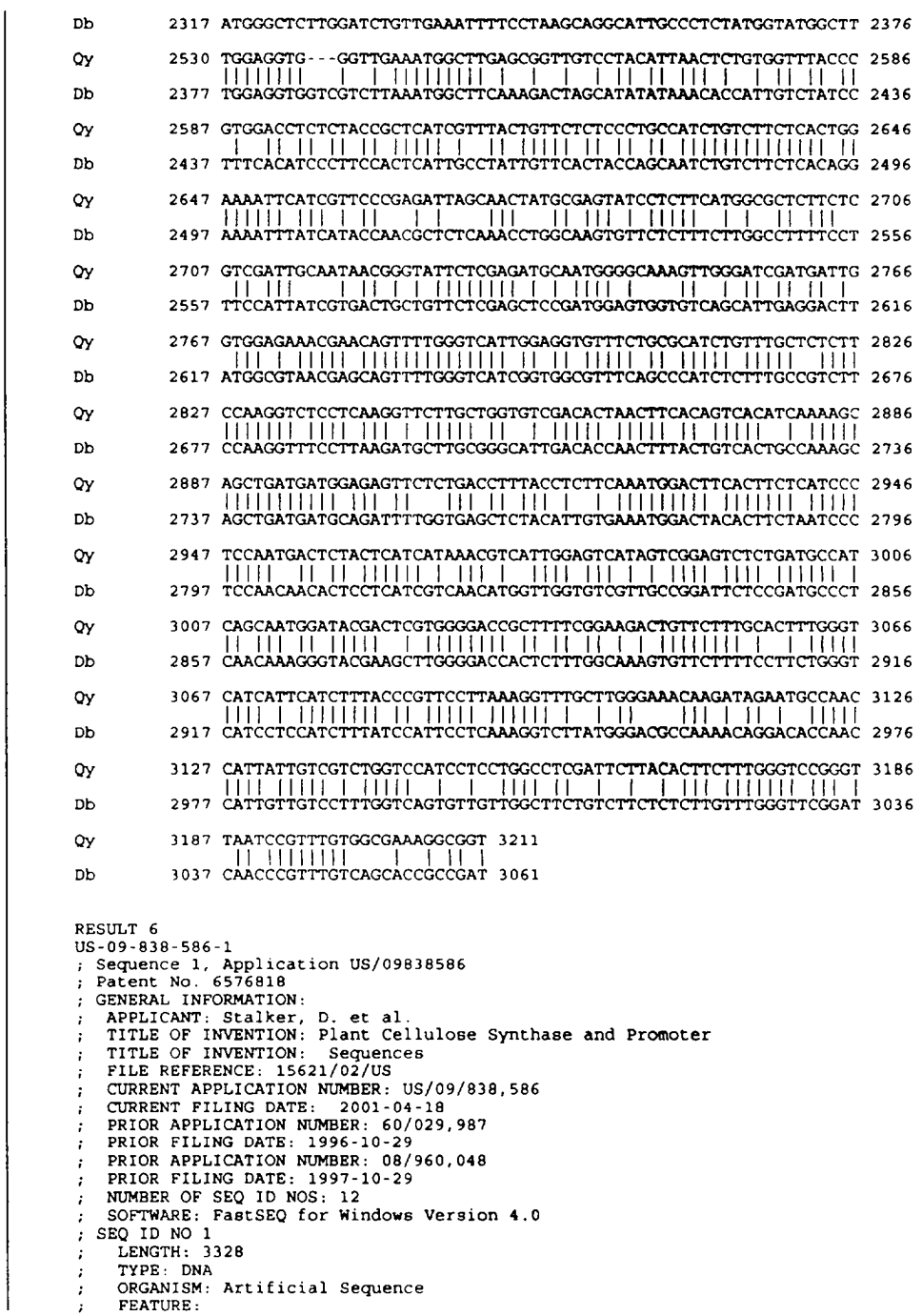
Qy      1210 TTCTGAGACCGCTGAATTCGCAAGGAAATGGGTTCCTTTCTGCAAGAAATATTGTATTGA 1269
Db      1087 AGTAGAAACAGCCGACTTTGCAAGAAAGTGGGTTCCATTCTGCAAAAAATTTTCCATTGA 1146

Qy      1270 GCCTCGTGCTCCCGAATGGTATTTCTGCCATAAAATGGACTACTTGAAGAATAAAGTTCA 1329
Db      1147 ACCCCGGGCACCTGAGTTTTACTTCTCACAGAAGATTGATTACTTGAAAGATAAAGTGCA 1206

Qy      1330 TCCCGCATTTGTTAGGGAGCGGCGAGCCATGAAGAGAGATTATGAAGAATTCAAAGTAAA 1389
Db      1207 GCCCTCTTTTGTAAAAGAACGTAGAGCTATGAAAAGAGATTATGAAGAGTACAAAATTCG 1266

Qy      1390 GATCAATGCTTTAGTAGCAACAGCACAGAAAGTGCCTGAGGATGGTTGGACTATGCAAGA 1449
```

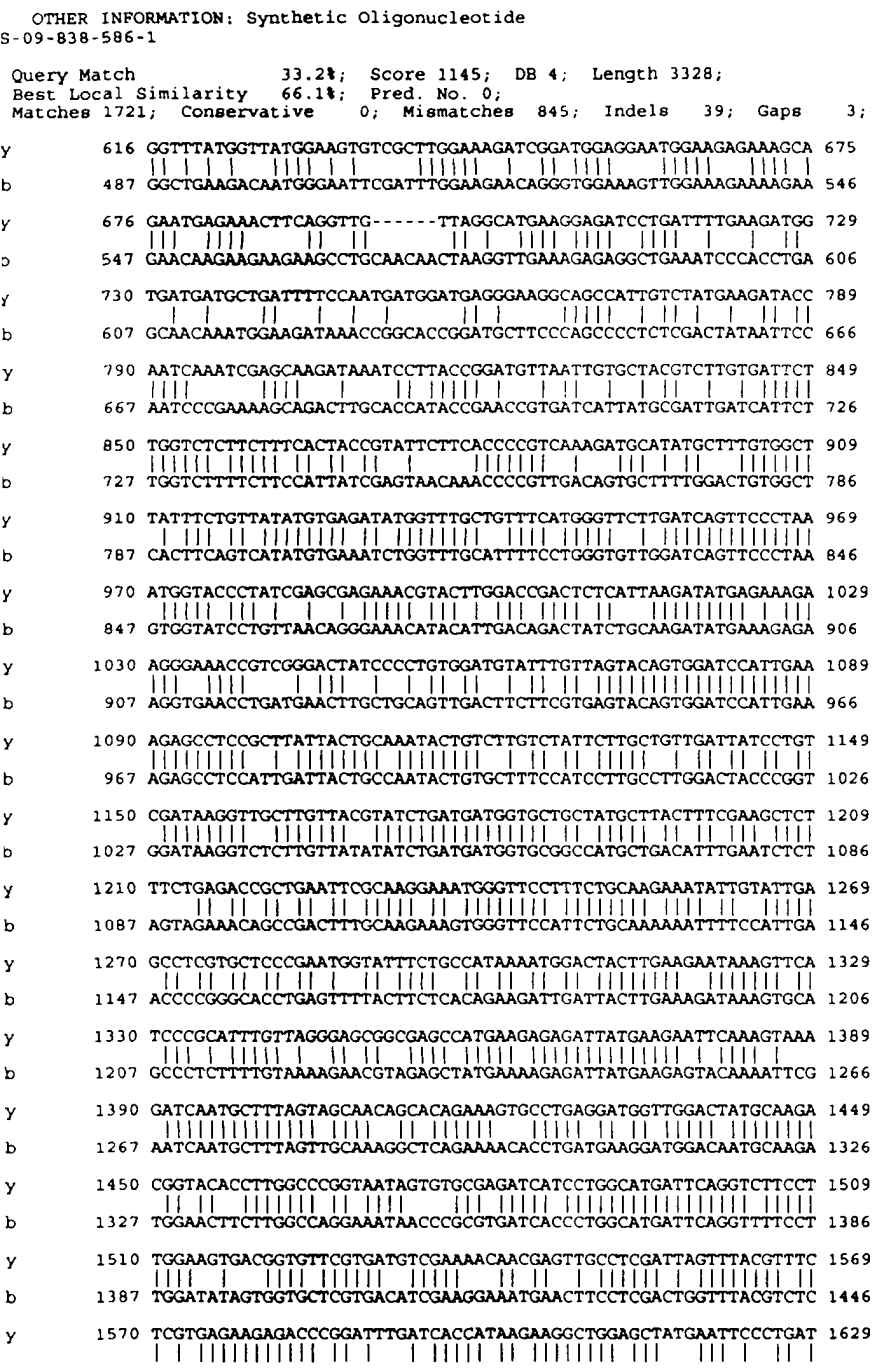


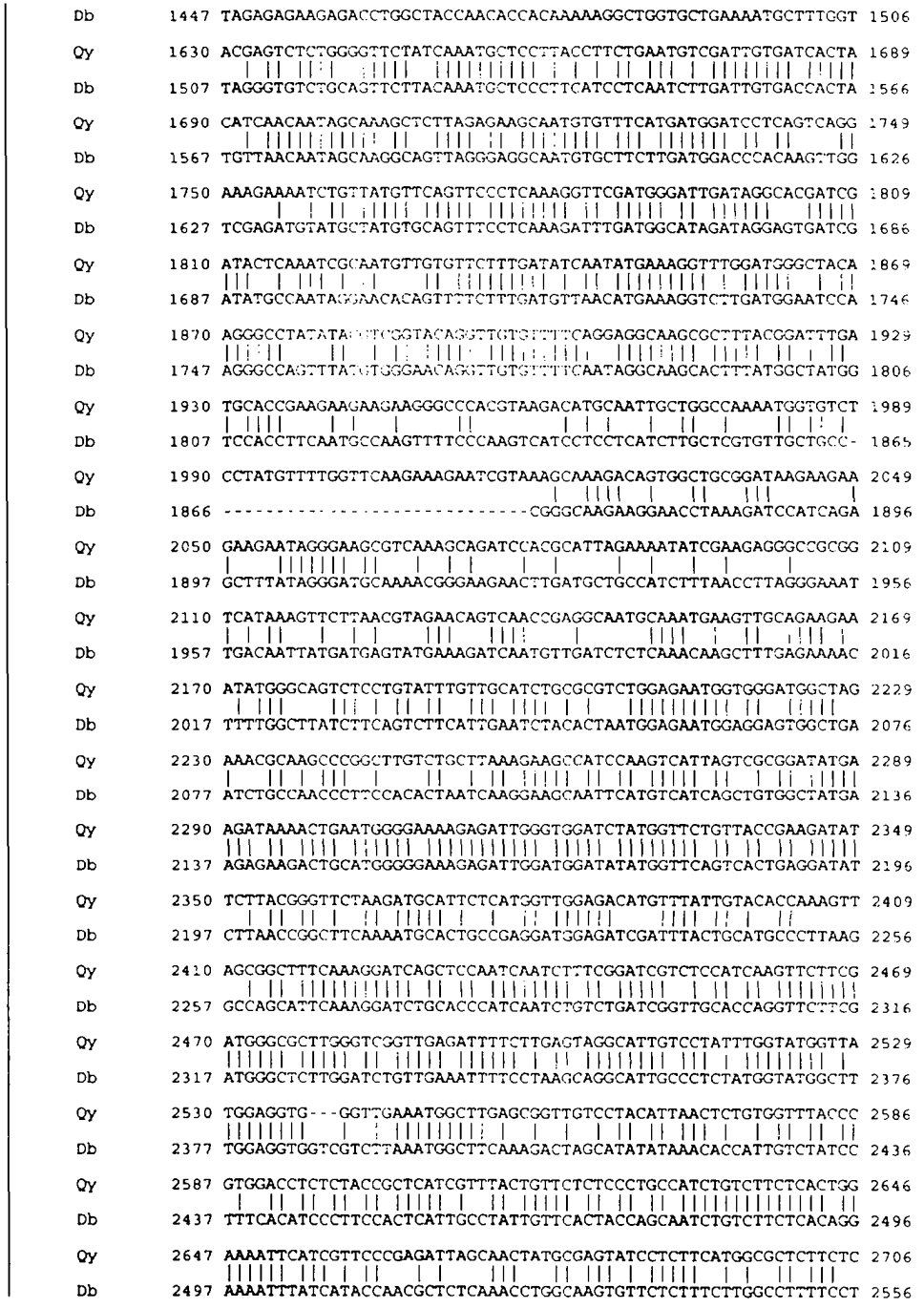


```
RESULT 6
US-09-838-586-1
; Sequence 1, Application US/09838586
; Patent No. 6576818
; GENERAL INFORMATION:
;  APPLICANT: Stalker, D. et al.
;  TITLE OF INVENTION: Plant Cellulose Synthase and Promoter
;  TITLE OF INVENTION:  Sequences
;  FILE REFERENCE: 15621/02/US
;  CURRENT APPLICATION NUMBER: US/09/838,586
;  CURRENT FILING DATE:  2001-04-18
;  PRIOR APPLICATION NUMBER: 60/029,987
;  PRIOR FILING DATE: 1996-10-29
;  PRIOR APPLICATION NUMBER: 08/960,048
;  PRIOR FILING DATE: 1997-10-29
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 3328
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
```

OTHER INFORMATION: Synthetic Oligonucleotide
S-09-838-586-1

Query Match 33.2%; Score 1145; DB 4; Length 3328;
Best Local Similarity 66.1%; Pred. No. 0;
Matches 1721; Conservative 0; Mismatches 845; Indels 39; Gaps 3;

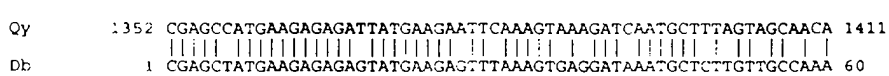


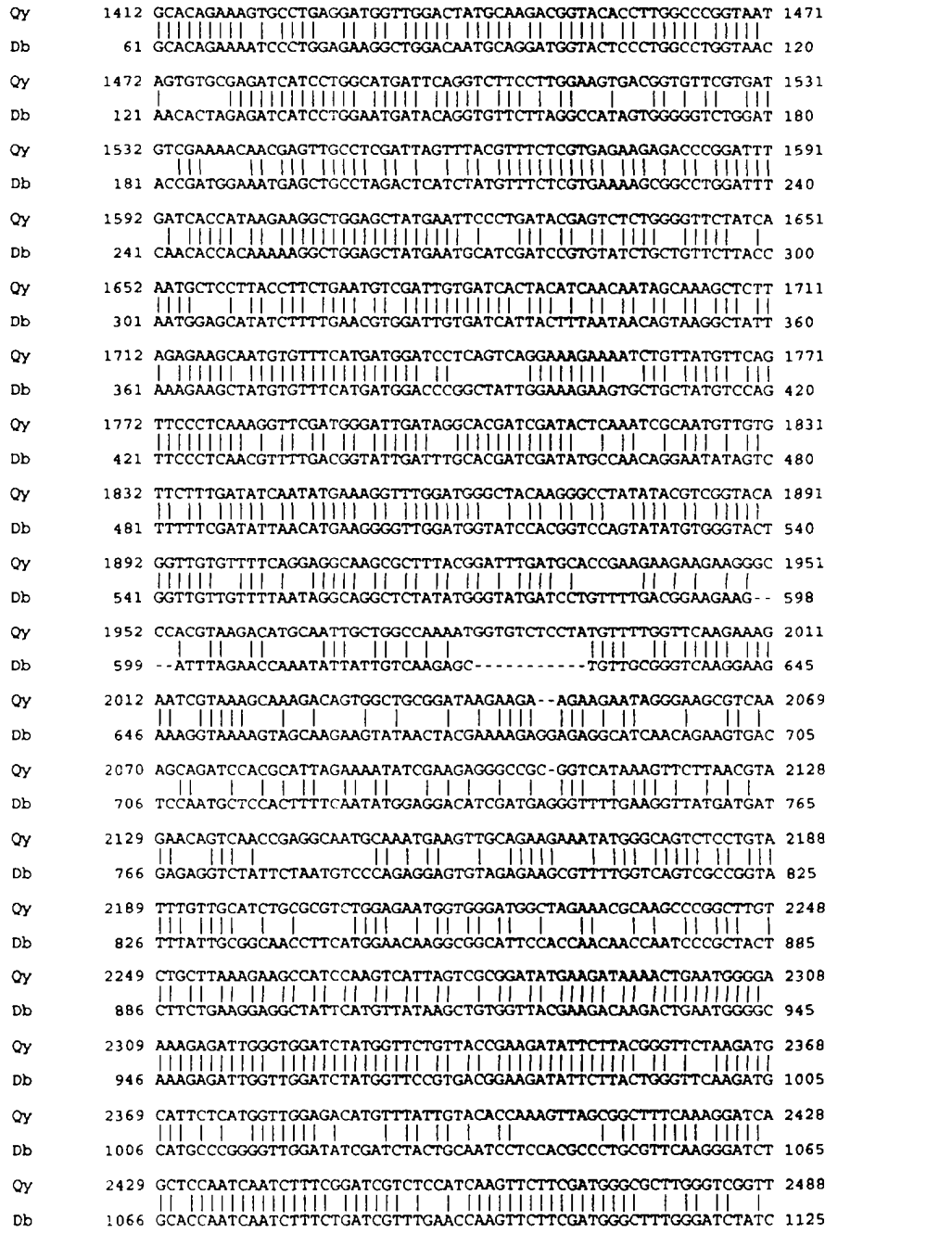


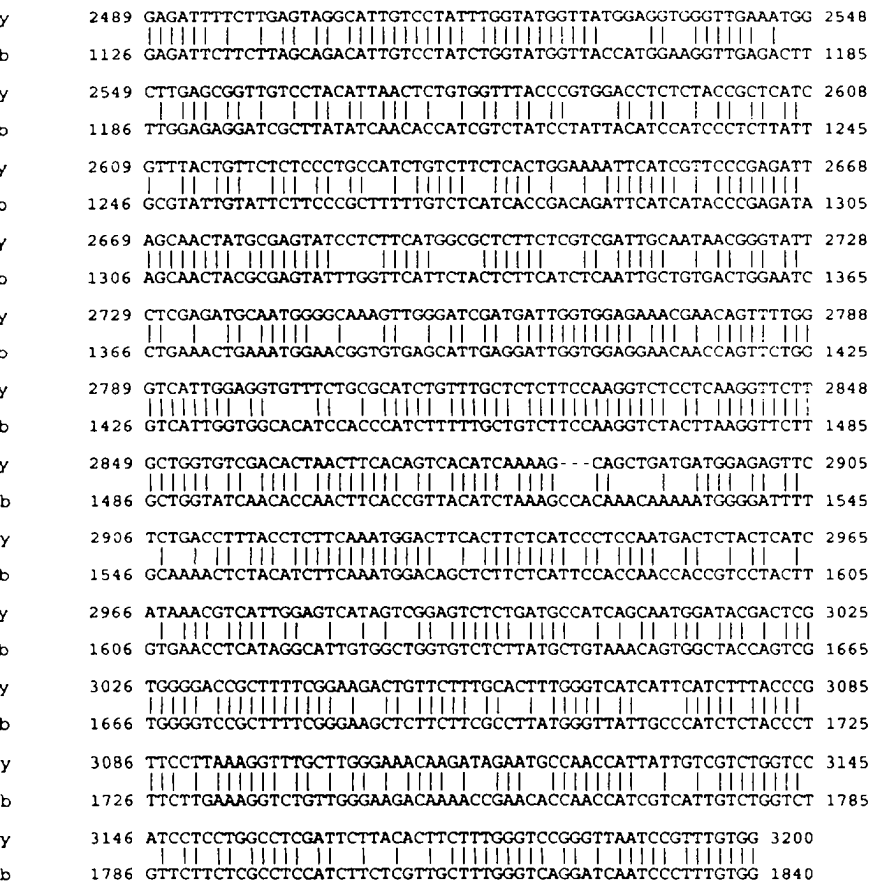


RESULT 7
US-09-221-013A-1
; Sequence 1, Application US/09221013A
; Patent No. 6495740
; GENERAL INFORMATION:
; APPLICANT: Arioli, Antonio
; APPLICANT: Williamson, Richard E.
; APPLICANT: Betzner, Andreas S.
; APPLICANT: Peng, Liangcai
; TITLE OF INVENTION: Manipulation of cellulose and/or Beta-1,4-glucan
; FILE REFERENCE: 96-98
; CURRENT APPLICATION NUMBER: US/09/221,013A
; CURRENT FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: PCT/AU97/00402
; PRIOR FILING DATE: 1997-06-24
; PRIOR APPLICATION NUMBER: AU PO0699
; PRIOR FILING DATE: 1996-06-27
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2248
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1887)
US-09-221-013A-1

Query Match 23.9%; Score 822.2; DB 4; Length 2248;
Best Local Similarity 67.2%; Pred. No. 1.5e-237;
Matches 1246; Conservative 0; Mismatches 588; Indels 21; Gaps 5;



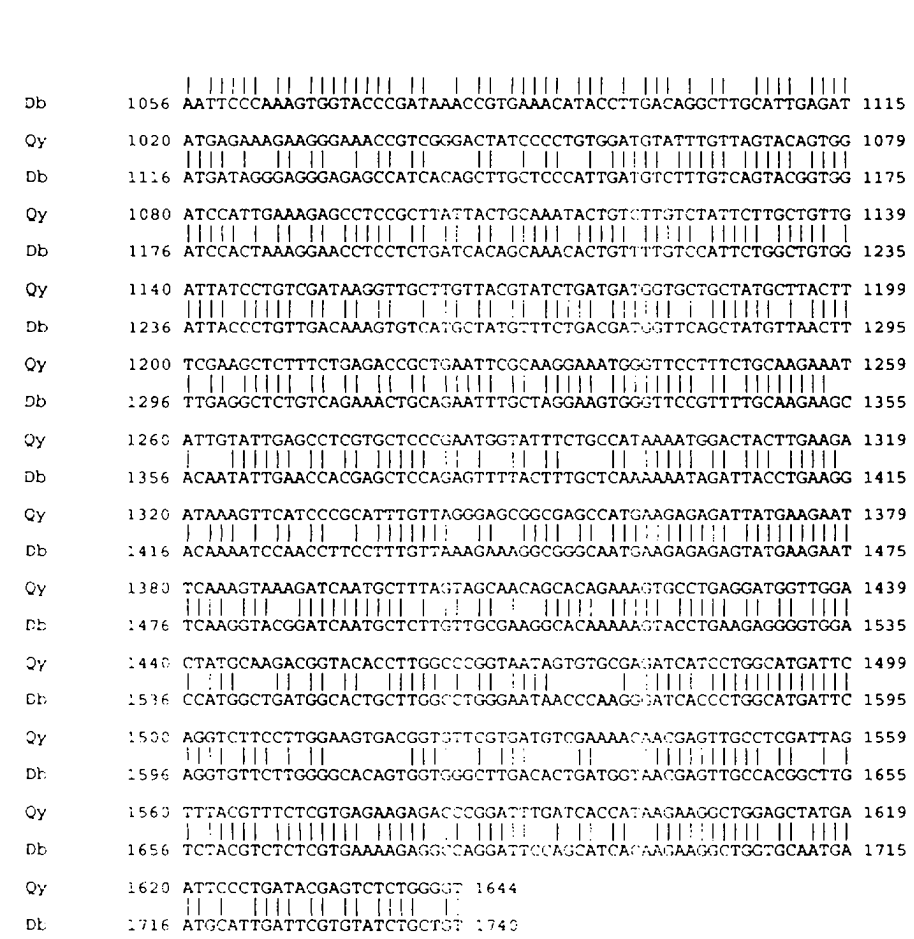




ESULT 8
S-09-221-013A-13
Sequence 13, Application US/09221013A
Patent No. 6495740
GENERAL INFORMATION:
APPLICANT: Arioli, Antonio
APPLICANT: Williamson, Richard E.
APPLICANT: Betzner, Andreas S.
APPLICANT: Peng, Liangcai
TITLE OF INVENTION: Manipulation of cellulose and/or Beta-1,4-glucan
FILE REFERENCE: 96-98
CURRENT APPLICATION NUMBER: US/09/221,013A
CURRENT FILING DATE: 1998-12-23
PRIOR APPLICATION NUMBER: PCT/AU97/00402
PRIOR FILING DATE: 1997-06-24
PRIOR APPLICATION NUMBER: AU PO0699
PRIOR FILING DATE: 1996-06-27
NUMBER OF SEQ ID NOS: 37
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 13
LENGTH: 1741
TYPE: DNA
ORGANISM: Oryza sativa
FEATURE:

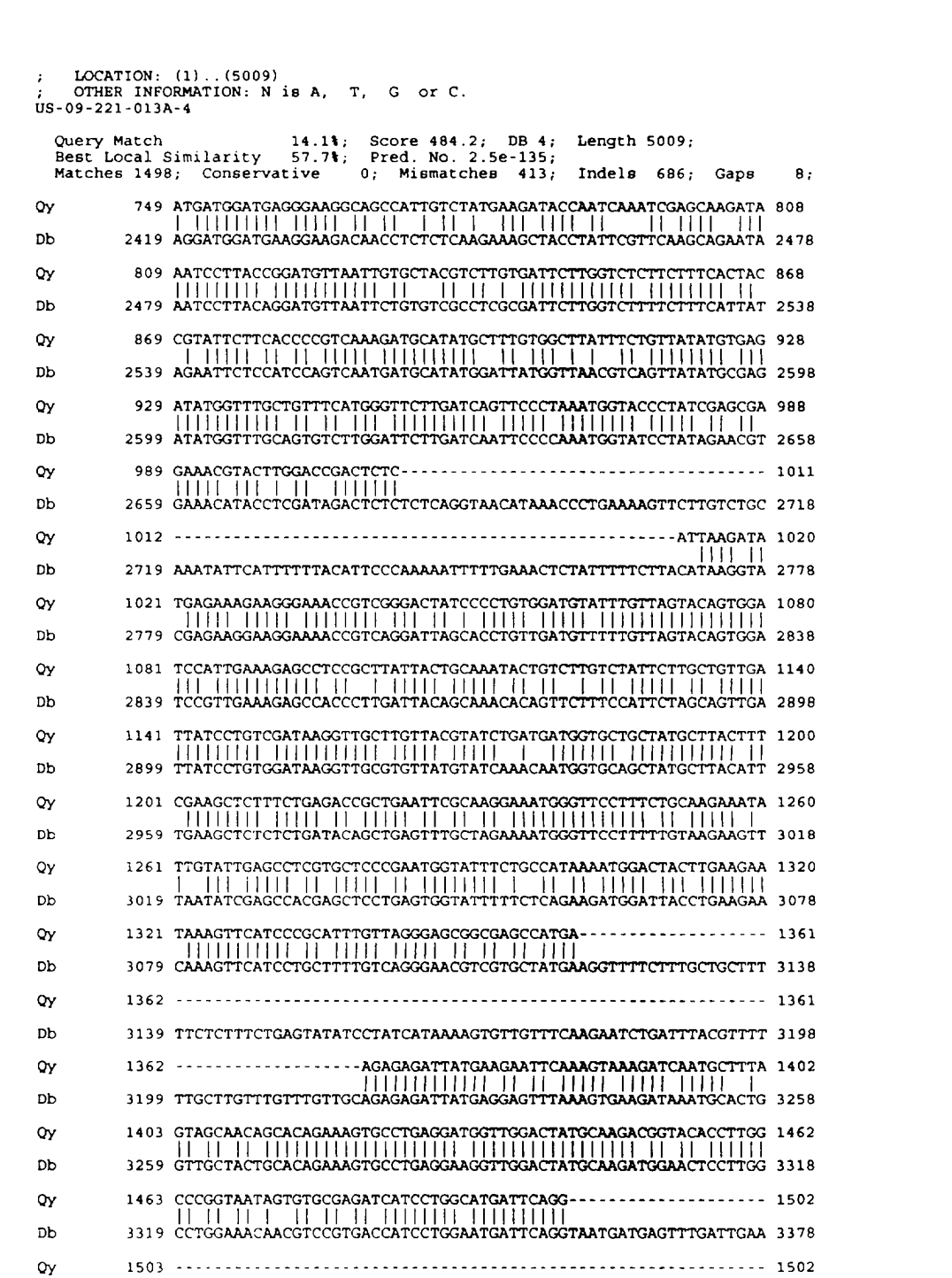
; NAME/KEY: CDS
; LOCATION: (101)...1741
US-09-221-013A-13

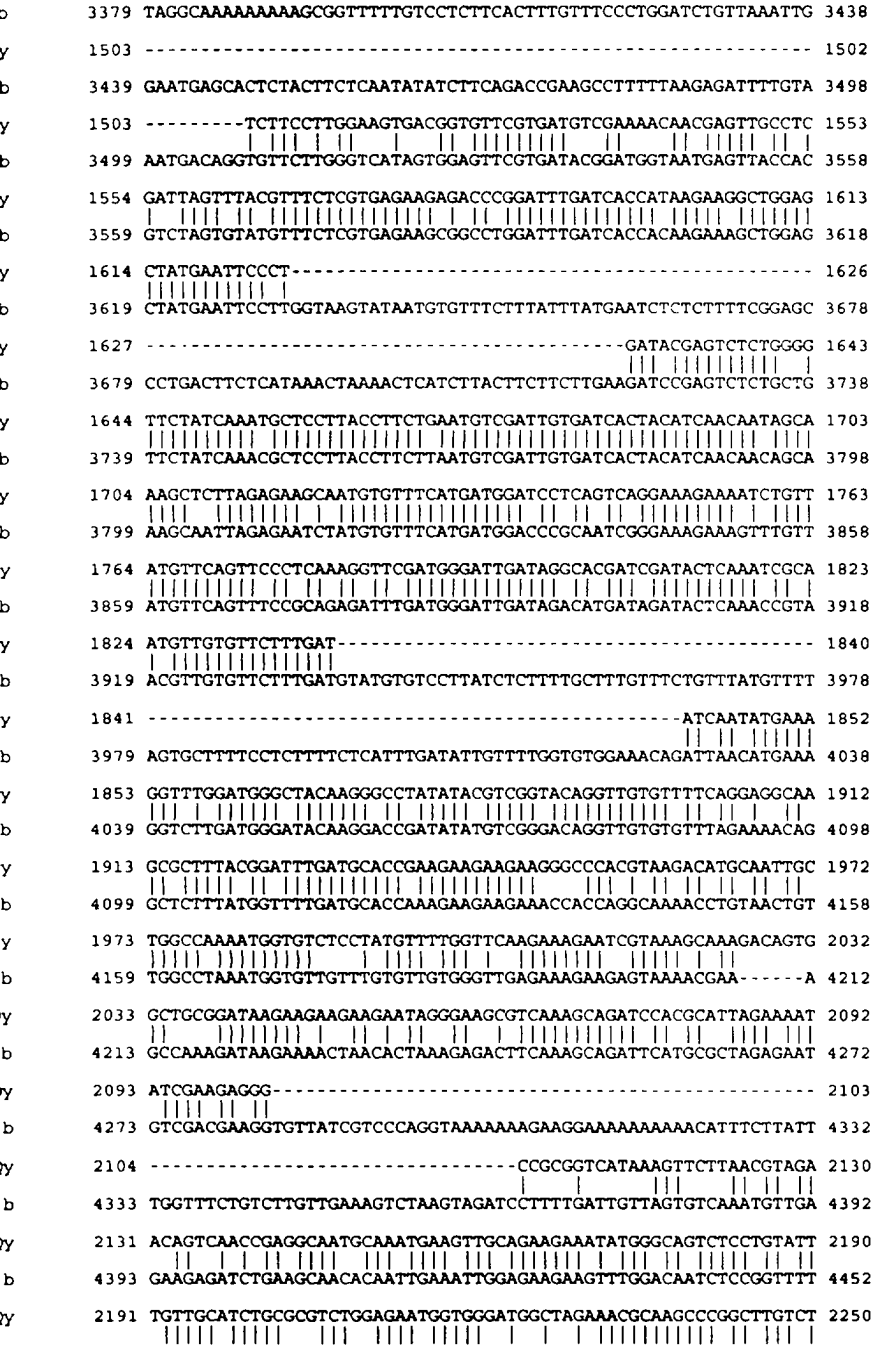
Query Match 15.9%; Score 546.2; DB 4; Length 1741;
Best Local Similarity 54.8%; Pred. No. 2.6e-154;
Matches 983; Conservative 0; Mismatches 638; Indels 24; Gaps 3;

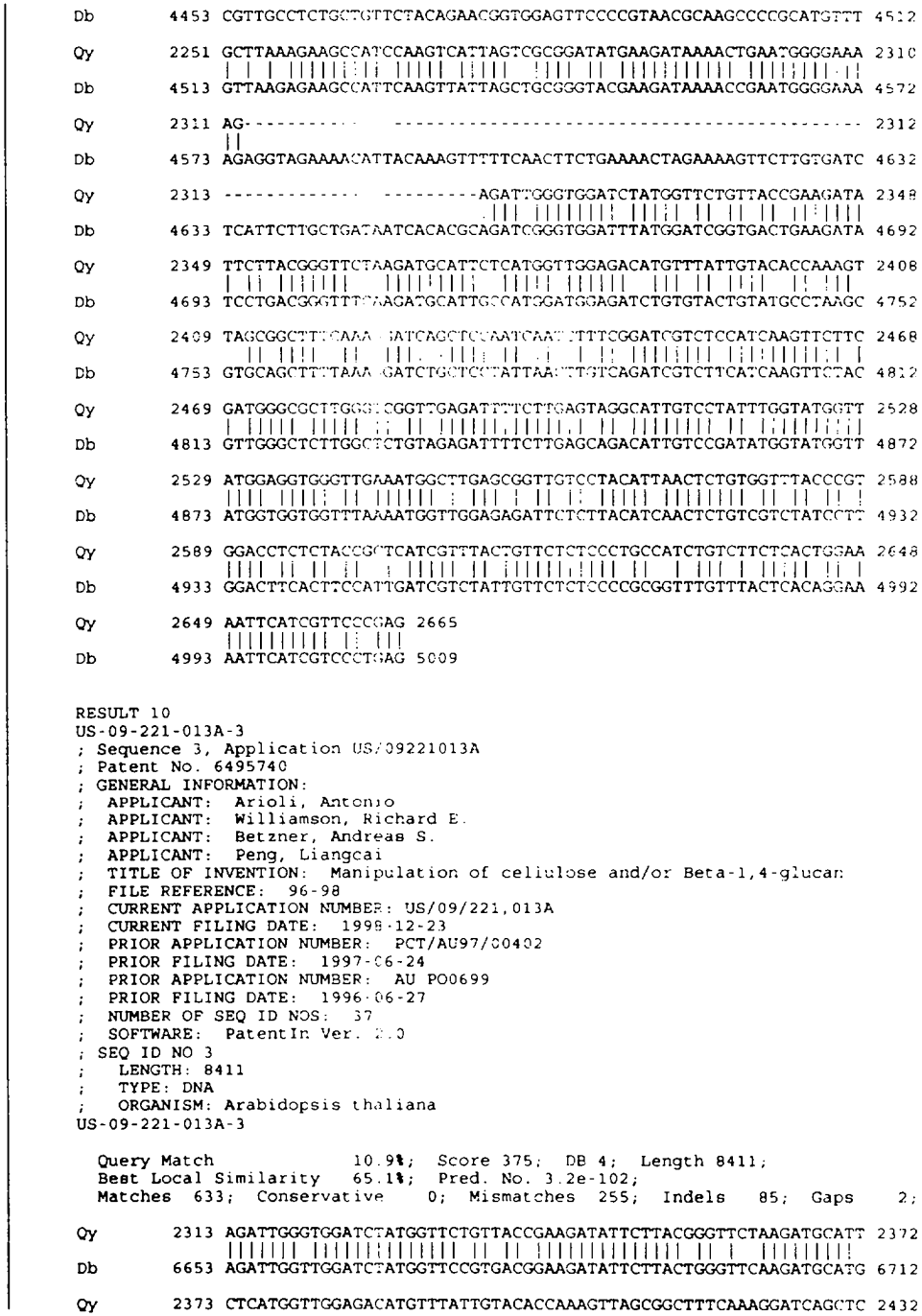


```
RESULT 9
US-09-221-013A-4
; Sequence 4, Application US/09221013A
; Patent No. 6495740
; GENERAL INFORMATION:
;  APPLICANT:  Arioli, Antonio
;  APPLICANT:  Williamson, Richard E.
;  APPLICANT:  Betzner, Andreas S.
;  APPLICANT:  Peng, Liangcai
;  TITLE OF INVENTION:  Manipulation of cellulose and/or Beta-1,4-glucan
;  FILE REFERENCE:  96-98
;  CURRENT APPLICATION NUMBER: US/09/221,013A
;  CURRENT FILING DATE:  1998-12-23
;  PRIOR APPLICATION NUMBER:  PCT/AU97/00402
;  PRIOR FILING DATE:  1997-06-24
;  PRIOR APPLICATION NUMBER:  AU PO0699
;  PRIOR FILING DATE:  1996-06-27
;  NUMBER OF SEQ ID NOS:  37
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 4
;   LENGTH: 5009
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(5009)
;   OTHER INFORMATION: N is A,  T,  G  or C.
US-09-221-013A-4

  Query Match            14.1%;  Score 484.2;  DB 4;  Length 5009;
  Best Local Similarity  57.7%;  Pred. No. 2.5e-135;
  Matches 1498;  Conservative    0;  Mismatches  413;  Indels  686;  Gaps    8;
```







RESULT 10
US-09-221-013A-3
; Sequence 3, Application US/09221013A
; Patent No. 6495740
; GENERAL INFORMATION:
; APPLICANT: Arioli, Antonio
; APPLICANT: Williamson, Richard E.
; APPLICANT: Betzner, Andreas S.
; APPLICANT: Peng, Liangcai
; TITLE OF INVENTION: Manipulation of cellulose and/or Beta-1,4-glucan
; FILE REFERENCE: 96-98
; CURRENT APPLICATION NUMBER: US/09/221,013A
; CURRENT FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: PCT/AU97/00402
; PRIOR FILING DATE: 1997-06-24
; PRIOR APPLICATION NUMBER: AU PO0699
; PRIOR FILING DATE: 1996-06-27
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 8411
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-221-013A-3

Query Match 10.9%; Score 375; DB 4; Length 8411;
Best Local Similarity 65.1%; Pred. No. 3.2e-102;
Matches 633; Conservative 0; Mismatches 255; Indels 85; Gaps 2;

```
Db       6713 CCCGGGGTTGGATATCGATCTACTGCAATCCTCCACGCCCTGCGTTCAAGGGATCTGCAC 6772

Qy       2433 CAATCAATCTTTCGGATCGTCTCCATCAAGTTCTTCGATGGGCGCTTGGGTCGGTTGAGA 2492

Db       6773 CAATCAATCTTTCTGATCGTTTGAACCAAGTTCTTCGATGGGCTTTGGGATCTATCGAGA 6832

Qy       2493 TTTTCTTGAGTAGGCATTGTCCTATTTGGTATGGTTATGGAGGTGGGTTGAAATGGCTTG 2552

Db       6833 TTCTTCTTAGCAGACATTGTCCTATCTGGTATGGTTACCATGGAAGGTTGAGACTTTTGG 6892

Qy       2553 AGCGGTTGTCCTACATTAACTCTGTGGTTTACCCGTGGACCTCTCTACCGCTCATCGTTT 2612

Db       6893 AGAGGATCGCTTATATCAACACCATCGTCTATCCTATTACATCCATCCCTCTTATTGCGT 6952

Qy       2613 ACTGTTCTCTCCCTGCCATCTGTCTTCTCACTGGAAAATTCATCGTTCCCG--------- 2663

Db       6953 ATTGTATTCTTCCCGCTTTTTGTCTCATCACCGACAGATTCATCATACCCGAGGTTTGTA 7012

Qy       2664 ------------------------------------------------------- 2663

Db       7013 AAACTGACCACACTGCTATTTACTATTTGAATCCCATTTTGTGAATGCATTTTTTTGTCA 7072

Qy       2664 ------------AGATTAGCAACTATGCGAGTATCCTCTTCATGGCGCTCTTCTCGTCG 2710

Db       7073 TCATCATTGTTGCAGATAAGCAACTACGCGAGTATTTGGTTCATTCTACTCTTCATCTCA 7132

Qy       2711 ATTGCAATAACGGGTATTCTCGAGATGCAATGGGGCAAAGTTGGGATCGATGATTGGTGG 2770

Db       7133 ATTGCTGTGACTGGAATCCTGGAGCTGAGATGGAGCGGTGTGAGCATTGAGGATTGGTGG 7192

Qy       2771 AGAAACGAACAGTTTTGGGTCATTGGAGGTGTTTCTGCGCATCTGTTTGCTCTCTTCCAA 2830

Db       7193 AGGAACGAGCAGTTCTGGGTCATTGGTGGCACATCCGCCCATCTTTTTGCTGTCTTCCAA 7252

Qy       2831 GGTCTCCTCAAGGTTCTTGCTGGTGTCGACACTAACTTCACAGTCACATCAAAAG---CA 2887

Db       7253 GGTCTACTTAAGGTTCTTGCTGGTATCGACACCAACTTCACCGTTACATCTAAAGCCACA 7312

Qy       2888 GCTGATGATGGAGAGTTCTCTGACCTTTACCTCTTCAAATGGACTTCACTTCTCATCCCT 2947

Db       7313 GACGAAGATGGGGATTTTGCAGAACTCTACATCTTCAAATGGACAGCTCTTCTCATTCCA 7372

Qy       2948 CCAATGACTCTACTCATCATAAACGTCATTGGAGTCATAGTCGGAGTCTCTGATGCCATC 3007

Db       7373 CCAACCACCGTCCTACTTGTGAACCTCATAGGCATTGTGGCTGGTGTCTCTTATGCTGTA 7432

Qy       3008 AGCAATGGATACGACTCGTGGGGACCGCTTTTCGGAAGACT-CTTCTTGCACTTTGGGTC 3067

Db       7433 AACAGTGGCTACCAGTCGTGGGGTCCGCTTTTCGGAAGCC-TTCTTGGCCTTATGGGTT 7492

Qy       3068 ATCATTCATCTTTACCCGTTCCTTAAAGGTTTGCTTGGGAAA-AAGATAGAATGCCAACC 3127

Db       7493 ATTGCCCATCTCTACCCTTTCTTGAAAGGTCTGTTGGGAAGACAAAACCGAACACCAACC 7552

Qy       3128 ATTATTGTCGTCTGGTCCATCCTCCTGGCCTCGATTCTTACACTTCTTTGGGTCCGGGTT 3187

Db       7553 ATCGTCATTGTCTGGTCTGTTCTTCTCGCCTCCATCTTCTCGTTGCTTTGGGTCAGGATC 7612

Qy       3188 AATCCGTTTGTGG 3200

Db       7613 AATCCCTTTGTGG 7625


RESULT 11
US-09-313-294A-6485
; Sequence 6485, Application US/09313294A
; Patent No. 6476212
; GENERAL INFORMATION:
;  APPLICANT:  Lalgudi, Raghunath V.
;  APPLICANT:  Ito, Laura Y.
;  APPLICANT:  Sherman, Bradley K.
;  TITLE OF INVENTION:  POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
;  FILE REFERENCE:  PL-0017 US
;  CURRENT APPLICATION NUMBER: US/09/313,294A
;  CURRENT FILING DATE:  1999-05-14
;  NUMBER OF SEQ ID NOS:  7600
;  SOFTWARE:  PERL Program
; SEQ ID NO 6485
;   LENGTH: 281
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. 6476212 700351882H1
US-09-313-294A-6485

  Query Match             4.5%;  Score 155.8;  DB 4;  Length 281;
  Best Local Similarity   72.4%;  Pred. No. 5.4e-37;
  Matches  202;  Conservative    0;  Mismatches   77;  Indels    0;  Gaps    0;

Qy       2566 CATTAACTCTGTGGTTTACCCGTGGACCTCTCTACCGCTCATCGTTTACTGTTCTCTCCC 2625

Db          1 CATCAACACCACCATCTACCCGCTCACGTCCCTCCCGCTCCTCATTTACTGTATCCTGCC 60

Qy       2626 TGCCATCTGTCTTCTCACTGGAAAATTCATCGTTCCCGAGATTAGCAACTATGCGAGTAT 2685

Db         61 TGCCATCTGCCTGCTCACGGGGAAGTTCATCATCCCAGAGATCAGCAACTTCGCTAGTAT 120

Qy       2686 CCTCTTCATGGCGCTCTTCTCGTCGATTGCAATAACGGGTATTCTCGAGATGCAATGGGG 2745

Db        121 CTGGTTCATCTCTCTCTTCATCTCGATCTTCGCCACGGGTATCCTGGAGATGAGGTGGAG 180

Qy       2746 CAAAGTTGGGATCGATGATTGGTGGAGAAACGAACAGTTTTGGGTCATTGGAGGTGTTTC 2805

Db        181 CGGCGTGGGCATCGACGAGTGGTGGAGGAACGAGCAGTTCTGGGTCATCGGAGGCATCTC 240

Qy       2806 TGCGCATCTGTTTGCTCTCTTCCAAGGTCTCCTCAAGGT 2844

Db        241 CGCCCACCTCTTCGCCGTCTTCCAGGGCCTCCTCAAGGT 279


RESULT 12
US-09-313-294A-5519
; Sequence 5519, Application US/09313294A
; Patent No. 6476212
; GENERAL INFORMATION:
;  APPLICANT:  Lalgudi, Raghunath V.
;  APPLICANT:  Ito, Laura Y.
;  APPLICANT:  Sherman, Bradley K.
;  TITLE OF INVENTION:  POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
;  FILE REFERENCE:  PL-0017 US
;  CURRENT APPLICATION NUMBER: US/09/313,294A
;  CURRENT FILING DATE:  1999-05-14
;  NUMBER OF SEQ ID NOS:  7600
;  SOFTWARE:  PERL Program
; SEQ ID NO 5519
;   LENGTH: 291
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. 6476212 700350285H1
;   NAME/KEY: unsure
;   LOCATION: 6, 40, 54, 65, 67, 200, 221, 230, 238, 261
;   OTHER INFORMATION: a, t, c, g, or other
US-09-313-294A-5519

  Query Match             4.0%;  Score 137.8;  DB 4;  Length 291;
  Best Local Similarity   68.3%;  Pred. No. 1.5e-31;
  Matches  198;  Conservative    0;  Mismatches   91;  Indels    1;  Gaps    1;

Qy       2373 CTCATGGTTGGAGACATGTTTATTGTACACCAAAGTTAGCGGCTTTCAAAGGATCAGCTC 2432

Db          1 CTCGANGCTGGCGATCAATCTACTGCATGCCTAAGCGACNAGCTTTCAAGGGANCTGCTC 60
```

```
Qy       2433 CAAT-CAATCTTTCGGATCGTCTCCATCAAGTTCTTCGATGGGCGCTTGGGTCGGTTGAG 2491
Db         61 CTATNCNACCTTTCGGATCGTTTGAATCAAGTGCTTCGGTGGGCTCTTGGTTCCATTGAA 120

Qy       2492 ATTTTCTTGAGTAGGCATTGTCCTATTTGGTATGGTTATGGAGGTGGGTTGAAATGGCTT 2551
Db        121 ATTCTTTTCAGCAGGCATTGTCCCATATGGTATGGCTATGGAGGTCGGCTTAAATTCCTG 180

Qy       2552 GAGCGGTTGTCCTACATTAACTCTGTGGTTTACCCGTGGACCTCTCTACCGCTCATCGTT 2611
Db        181 GAGAGATTTGCTTATATCANCACAACAATTTATCCACTCANATCAATCCNGCTCCTCNTG 240

Qy       2612 TACTGTTCTCTCCCTGCCATCTGTCTTCTCACTGGAAAATTCATCGTTCC 2661
Db        241 TACTGCATATTGCCAGCAGTNTGTCTTCTCACTGGGAAGTTCATCATCCC 290


ESULT 13
S-09-313-294A-849
 Sequence 849, Application US/09313294A
 Patent No. 6476212
 GENERAL INFORMATION:
  APPLICANT:  Lalgudi, Raghunath V.
  APPLICANT:  Ito, Laura Y.
  APPLICANT:  Sherman, Bradley K.
  TITLE OF INVENTION:  POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
  FILE REFERENCE:  PL-0017 US
  CURRENT APPLICATION NUMBER: US/09/313,294A
  CURRENT FILING DATE:  1999-05-14
  NUMBER OF SEQ ID NOS:  7600
  SOFTWARE:  PERL Program
 SEQ ID NO 849
    LENGTH: 268
    TYPE: DNA
    ORGANISM: Zea mays
    FEATURE:
    NAME/KEY: misc_feature
    OTHER INFORMATION: Incyte ID No. 6476212 700549940H1
S-09-313-294A-849

 Query Match             4.0%;  Score 137.4;  DB 4;  Length 268;
 Best Local Similarity   69.7%;  Pred. No. 1.9e-31;
 Matches  186;  Conservative    0;  Mismatches   81;  Indels    0;  Gaps    0;

Qy       2695 GGCGCTCTTCTCGTCGATTGCAATAACGGGTATTCTCGAGATGCAATGGGGCAAAGTTGG 2754
Db          1 GTCACTTTTTATCTGCATTTTTGCTACGAGCATCCTAGAAATGAGATGGAGTGGTGTTGG 60

Qy       2755 GATCGATGATTGGTGGAGAAACGAACAGTTTTGGGTCATTGGAGGTGTTTCTGCGCATCT 2814
Db         61 AATTGATGACTGGTGGAGGAATGAGCAGTTCTGGGTCATTGGAGGTGTGTCCTCACACCT 120

Qy       2815 GTTTGCTCTCTTCCAAGGTCTCCTCAAGGTTCTTGCTGGTGTCGACACTAACTTCACAGT 2874
Db        121 CTTTGCTGTGTTCCAGGGACTTCTCAAGGTCATAGCTGGTGTTGATACAAGCTTCACCGT 180

Qy       2875 CACATCAAAAGCAGCTGATGATGGAGAGTTCTCTGACCTTTACCTCTTCAAATGGACTTC 2934
Db        181 GACATCAAAGGGTGGAGATGATGAGGAGTTCTCAGAGCTATATACATTCAAATGGACTAC 240

Qy       2935 ACTTCTCATCCCTCCAATGACTCTACT 2961
Db        241 CTTATTGATACCTCCTACCACCTTGCT 267


ESULT 14
S-09-313-294A-4753
 Sequence 4753, Application US/09313294A
 Patent No. 6476212
 GENERAL INFORMATION:
  APPLICANT:  Lalgudi, Raghunath V.
  APPLICANT:  Ito, Laura Y.
  APPLICANT:  Sherman, Bradley K.
  TITLE OF INVENTION:  POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
  FILE REFERENCE:  PL-0017 US
  CURRENT APPLICATION NUMBER: US/09/313,294A
  CURRENT FILING DATE:  1999-05-14
  NUMBER OF SEQ ID NOS:  7600
  SOFTWARE:  PERL Program
 SEQ ID NO 4753
    LENGTH: 300
    TYPE: DNA
    ORGANISM: Zea mays
    FEATURE:
    NAME/KEY: misc_feature
    OTHER INFORMATION: Incyte ID No. 6476212 700348982H1
    NAME/KEY: unsure
    LOCATION: 64, 189
    OTHER INFORMATION: a, t, c, g, or other
US-09-313-294A-4753

  Query Match             4.0%;  Score 132.4;  DB 4;  Length 300;
  Best Local Similarity   69.7%;  Pred. No. 6.5e-30;
  Matches  199;  Conservative    0;  Mismatches   93;  Indels    2;  Gaps    1;

Qy       2664 AGATTAGCAACTATGCGAGTATCCTCTTCATGGCGCTCTTCTCGTCGATTGCAATAACGG 2723
Db          3 AGATCAGCAACTTCGCCAGCATCTGGTTCATCTCCCTCTTCATCTCGATCTTCGCCACGG 62

Qy       2724 GTATTCTCGAGATGCAATGGGGCAAAGTTGGGATCGATGATTGGTGGAGAAACGAACAGT 2783
Db         63 GNATCCTGGAGATGAGGTGGAGCGGGGTGGGCATCGACGAGTGGTGGAGGAACGAGCAGT 122

Qy       2784 TTTGGGTCATTGGAGGTGTTTCTGCGCATCTGTTTGCTCTCTTCCAAGGTCTCCTCAAGG 2843
Db        123 TCTGGGTGATCGGGGGCATCTCCGCGCACCTCTTCGCCGTGTTCCAGGGCCTGCTCAAGG 182

Qy       2844 TTCTTGCTGGTGTCGACACTAACTTCACAGTCACATCAAAAGCAGCTGATGATGGAGAGT 2903
Db        183 TGCTGGNCGGCATCGACACCAACTTCACCGTCACCTCCAAGGCCTCGGACGAGGACGGCG 242

Qy       2904 TCT--CTGACCTTTACCTCTTCAAATGGACTTCACTTCTCATCCCTCCAATGAC 2955
Db        243 ACTTCGCGGACTGTACATGTTCAAGTGGACGACGCTCCTTATCCCGCCCACCAC 296


RESULT 15
US-09-313-294A-4696
; Sequence 4696, Application US/09313294A
; Patent No. 6476212
; GENERAL INFORMATION:
;  APPLICANT:  Lalgudi, Raghunath V.
;  APPLICANT:  Ito, Laura Y.
;  APPLICANT:  Sherman, Bradley K.
;  TITLE OF INVENTION:  POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
;  FILE REFERENCE:  PL-0017 US
;  CURRENT APPLICATION NUMBER: US/09/313,294A
;  CURRENT FILING DATE:  1999-05-14
;  NUMBER OF SEQ ID NOS:  7600
;  SOFTWARE:  PERL Program
; SEQ ID NO 4696
;   LENGTH: 286
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. 6476212 700348878H1
;   NAME/KEY: unsure
;   LOCATION: 91-92, 182, 252
;   OTHER INFORMATION: a, t, c, g, or other
US-09-313-294A-4696

  Query Match             3.6%;  Score 123;  DB 4;  Length 286;
  Best Local Similarity   66.4%;  Pred. No. 4.4e-27;
  Matches  188;  Conservative    0;  Mismatches   94;  Indels    1;  Gaps    1;
```

```
Qy       2664 AGATTAGCAACTATGCGAGTATCCTCTTCATGGCGCTCTTCTCGTCGATTGCAATAACGG 2723
              |||| |||||||  || || |||  |||||  | ||||||  |||||      ||||
Db          4 AGATCAGCAACTTCGCCAGCATCTGGTTCATCTCCCTCTTCATCTCGATCTTCGCCACGG 63

Qy       2724 GTATTCTCGAGATGCAATGGGGCAAAGTTGGGATCGATGATTGGTGGAGAAACGAACAGT 2783
              | || || ||||||   ||| ||   |   ; ||||| || |||||||| ||||| ||||
Db         64 GCATCCTGGAGATGAGGTGGAGCGGGGNNCGCATCGACGAGTGGTGGAGGAACGAGCAGT 123

Qy       2784 TTTGGGTCATTGGAGGTGTTTCTGCGCATCTGTTTGCTCTCTTTCCAAGGTCTCCTCAAGG 2843
              | ||||| || || ||  | || ||||| |' || ||  | |||| || || ||||| |
Db        124 TCTGGGTGATCGGGGGCATCTCCGCGCACCTCTTCGCCGTGTTCCAGGGCCTGCTCAANG 183

Qy       2844 TTCTTGCTGGTGTCGACACTAACTTCACAGTCACATCAAAAGC AGCTGATGATGGAGAG 2902
              | || || ||  |||||| |||;||| !'||| || || :|  |     |   |
Db        184 TGCTGGCCGGCATCGACACCAACTTCACCGTCACCTCCAAGGCTCGGACGAGGACGGCGA 243

Qy       2903 TTCTCTGACCTTTACCTCTTCAAATGGACTTCACTTCTCATCC 2945
              ||| | ||  | || | ||||| |||!  | || || |:||
Db        244 TTCGCGGANTGTAACATGTTCAAGTGGACGACCCTCCTAATCC 286
```

Search completed: October 30, 2003, 16:31:53
Job time : 194.589 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 30, 2003, 05:34:29 ; Search time 837.27 Seconds
(without alignments)
11103.788 Million cell updates/sec

Title: US-09-980-043-5
Perfect score: 3444
Sequence: 1 gcggccgcggttaatcgccg.....gtgttgtagatagcggccgc 3444

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3444	100.0	3444	22	AAC83798	Arabidopsis thalia
2	3211.2	93.2	3255	24	ABZ14725	Arabidopsis thalia
3	1944	56.4	3255	24	ABZ12754	Arabidopsis thalia
4	1934.4	56.2	3828	19	AAV06566	Arabidopsis cellul
5	1489.8	43.3	3936	21	AAZ58266	Corn cellulose syn
6	1489.8	43.3	3969	21	AAZ99503	DNA encoding a mai
7	1489.8	43.3	3969	21	AAZ99518	DNA encoding a mai
8	1489	43.2	3776	21	AAZ58263	Corn cellulose syn
9	1464.8	42.5	3568	21	AAZ99491	DNA encoding a mai
10	1458.4	42.3	2890	21	AAZ58270	Soybean cellulose
11	1416.4	41.1	3786	21	AAZ58265	Corn cellulose syn
12	1416.4	41.1	3813	21	AAZ99509	DNA encoding a mai
13	1416.4	41.1	3813	21	AAZ99524	DNA encoding a mai
14	1294.8	37.6	3603	19	AAV06565	Arabidopsis cellul
15	1294.8	37.6	3673	19	AAV06568	Arabidopsis cellul
16	1277.6	37.1	3851	21	AAA67114	Pinus radiata cell
17	1257.4	36.5	3198	21	AAC49550	Arabidopsis thalia
18	1252.6	36.4	3614	19	AAV06567	Arabidopsis cellul
19	1212.8	35.2	3725	21	AAZ99500	DNA encoding a mai
20	1212.8	35.2	3725	21	AAZ99506	DNA encoding a mai
21	1212.8	35.2	3725	21	AAZ99521	DNA encoding a mai
22	1205.8	35.0	3517	21	AAZ58268	Soybean cellulose
23	1205.2	35.0	3753	21	AAZ99515	DNA encoding a mai
24	1205.2	35.0	3753	21	AAZ99530	DNA encoding a mai
25	1205.2	35.0	3780	21	AAZ99497	DNA encoding a mai
26	1204.4	35.0	3704	21	AAZ99533	DNA encoding a mai
27	1187	34.5	3311	19	AAV08373	Cellulose synthase
28	1168.4	33.9	3799	24	AAS16458	Corn cDNA encoding
29	1160	33.7	3746	21	AAZ99512	DNA encoding a mai
30	1160	33.7	3746	21	AAZ99527	DNA encoding a mai
31	1160	33.7	3773	21	AAZ99494	DNA encoding a mai
32	1147.8	33.3	3207	19	AAV08372	Cellulose synthase
33	1143.4	33.2	3328	19	AAV34432	Cotton cellulose s
34	1137.4	33.0	2830	24	AAS16455	Corn cDNA encoding
35	1086	31.5	3747	21	AAA67145	Eucalyptus grandis
36	1039	30.2	3232	22	AAC65448	Populus tremuloide
37	963.2	28.0	2306	19	AAV34433	Cotton cellulose s
38	822.2	23.9	2248	19	AAV06562	Arabidopsis cellul
39	806.8	23.4	2055	21	AAA67111	Eucalyptus grandis
40	802.4	23.3	2033	19	AAV08381	Cellulose synthase
41	795.6	23.1	2125	21	AAZ58269	Soybean cellulose
42	698	20.3	1860	21	AAA67081	Eucalyptus grandis
43	649	18.8	1734	21	AAZ58271	Wheat cellulose sy
44	636	18.5	1407	21	AAA67086	Eucalyptus grandis
45	579.8	16.8	1401	21	AAC44966	Arabidopsis thalia

ALIGNMENTS

RESULT 1
AAC83798
ID AAC83798 standard; cDNA; 3444 BP.
XX
AC AAC83798;
XX
DT 07-MAR-2001 (first entry)
XX
DE Arabidopsis thaliana cellulose synthase CelA cDNA.
XX
KW Arabidopsis thaliana; aspen; CelA; cellulose synthase; wood;
KW mechanical stress response element; MSRE; cellulose biosynthesis; ss.
XX
OS Arabidopsis thaliana.
XX
PN WO200071670-A2.
XX
PD 30-NOV-2000.
XX
PF 18-MAY-2000; 2000WO-US13637.
XX
PR 21-MAY-1999; 99US-0135280.
XX
PA (UNMT) UNIV MICHIGAN TECHNOLOGICAL.
XX
PI Chiang VL, Wu L, Chandrashekhar JP, Carraway DT;
XX
DR WPI; 2001-032025/04.
XX
PT Novel polynucleotide encoding cellulose synthase useful for enhancing
PT cellulose biosynthesis, modifying lignin biosynthesis, and for

T accelerating growth of plants and for upregulating cellulose synthase
T levels -
X
S Claim 8; Fig 7; 65pp; English.
X
C The present sequence is given in a specification relating to
C polynucleotides encoding cellulose synthase enzymes from Populus
C tremuloides and Arabidopsis. The cellulose synthases are designated
C PtCelA and AraxCelA respectively. The polynucleotides may be introduced
C into plants where they are useful for altering the growth of a plant.
C Vectors comprising a cellulose synthase promoter, or a fragment of the
C promoter containing one or more mechanical stress response elements
C (MSRE), operably linked to a cellulose synthase gene are useful for
C causing stress-induced gene expression in a plant cell. A cDNA encoding a
C protein that binds to a positive MSRE may be delivered into the plant to
C increase expression of cellulose synthase. A cDNA in an antisense
C orientation may be delivered to reduce expression of cellulose synthase.
C The polynucleotides enable the development of transgenic tree varieties
C having increased cellulose content, decreased lignin content and
C therefore improved wood fibre characteristics. The cellulose synthase
C promoters are useful for synthesising greater quantities of
C high-crystalline cellulose in plant and preferably in trees, which
C permits an increase in proportion of cellulose in transgenic plants,
C greater strength of juvenile wood and fiber and acceleration of overall
C growth rate.
C Note: This nucleotide sequence is given in the specification as
C SEQ ID NO: 5. SEQ ID NO: 5 is referred to in claim 8 but it is described
C as a polypeptide sequence.
X
Q Sequence 3444 BP; 903 A; 664 C; 860 G; 1017 T; 0 other;

Query Match 100.0%; Score 3444; DB 22; Length 3444;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3444; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GCGGCCGCGGTTAATCGCCGGTTCTCACAACAGGAATGAGTTTGTCCTCATTAATGCCGA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GCGGCCGCGGTTAATCGCCGGTTCTCACAACAGGAATGAGTTTGTCCTCATTAATGCCGA 60

Qy         61 TGAGAATGCCCGAATAAGATCAGTCCAAGAGCTGAGTGGACAGACATGTCAAATCTGCAG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TGAGAATGCCCGAATAAGATCAGTCCAAGAGCTGAGTGGACAGACATGTCAAATCTGCAG 120

Qy        121 AGATGAGATCGAATTGACTGTTGATGGAGAACCGTTTGTGGCATGTAACGAATGTGCATT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AGATGAGATCGAATTGACTGTTGATGGAGAACCGTTTGTGGCATGTAACGAATGTGCATT 180

Qy        181 CCCTGTGTGTAGACCTTGCTATGAGTACGAAAGACGAGAAGGCAATCAAGCTTGTCCACA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCCTGTGTGTAGACCTTGCTATGAGTACGAAAGACGAGAAGGCAATCAAGCTTGTCCACA 240

Qy        241 GTGCAAAACCCGTTTCAAACGTCTTAAAGGAAGTCCAAGAGTTGAAGGTGATGAAGAGGA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTGCAAAACCCGTTTCAAACGTCTTAAAGGAAGTCCAAGAGTTGAAGGTGATGAAGAGGA 300

Qy        301 AGATGACATTGATGATTTAGACAATGAGTTTGAGTATGGAAATAATGGGATTGGATTTGA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGATGACATTGATGATTTAGACAATGAGTTTGAGTATGGAAATAATGGGATTGGATTTGA 360

Qy        361 TCAGGTTTCTGAAGGTATGTCAATCTCTCGTCGCAACTCCGGTTTCCCACAATCTGATTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TCAGGTTTCTGAAGGTATGTCAATCTCTCGTCGCAACTCCGGTTTCCCACAATCTGATTT 420

Qy        421 GGATTCAGCTCCACCTGGCTCTCAGATTCCATTGCTGACTTACGGCGACGAGGACGTTGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGATTCAGCTCCACCTGGCTCTCAGATTCCATTGCTGACTTACGGCGACGAGGACGTTGA 480

Qy        481 GATTTCTTCTGATAGACATGCTCTTATTGTTCCTCCTTCACTTGGTGGTCATGGCAATAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GATTTCTTCTGATAGACATGCTCTTATTGTTCCTCCTTCACTTGGTGGTCATGGCAATAG 540

Qy        541 AGTTCATCCTGTTTCTCTTTCTGACCCGACCGTGGCTGCACATCGAAGGCTGATGGTACC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AGTTCATCCTGTTTCTCTTTCTGACCCGACCGTGGCTGCACATCGAAGGCTGATGGTACC 600

Qy        601 TCAGAAAGATCTTGCGGTTTATGGTTATGGAAGTGTCGCTTGGAAAGATCGGATGGAGGA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TCAGAAAGATCTTGCGGTTTATGGTTATGGAAGTGTCGCTTGGAAAGATCGGATGGAGGA 660

Qy        661 ATGGAAGAGAAAGCAGAATGAGAAACTTCAGGTTGTTAGGCATGAAGGAGATCCTGATTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ATGGAAGAGAAAGCAGAATGAGAAACTTCAGGTTGTTAGGCATGAAGGAGATCCTGATTT 720

Qy        721 TGAAGATGGTGATGATGCTGATTTTCCAATGATGGATGAGGGAAGGCAGCCATTGTCTAT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TGAAGATGGTGATGATGCTGATTTTCCAATGATGGATGAGGGAAGGCAGCCATTGTCTAT 780

Qy        781 GAAGATACCAATCAAATCGAGCAAGATAAATCCTTACCGGATGTTAATTGTGCTACGTCT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAAGATACCAATCAAATCGAGCAAGATAAATCCTTACCGGATGTTAATTGTGCTACGTCT 840

Qy        841 TGTGATTCTTGGTTTCTTTCTTCACTACCGTATCTTCACCCCGTCAAAGATGCATATGC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TGTGATTCTTGGTTTCTTTCTTCACTACCGTATCTTCACCCCGTCAAAGATGCATATGC 900

Qy        901 TTTGTGGCTTATTTCTGTTATATGTGAGATATGGTTTGCTGTTTCATGGGTTCTTGATCA 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTTGTGGCTTATTTCTGTTATATGTGAGATATGGTTTGCTGTTTCATGGGTTCTTGATCA 960

Qy        961 GTTCCCTAAATGGTACCCTATCGAGCGAGAAACGTACTTGGACCGACTCTCATTAAGATA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GTTCCCTAAATGGTACCCTATCGAGCGAGAAACGTACTTGGACCGACTCTCATTAAGATA 1020

Qy       1021 TGAGAAAGAAGGGAAACCGTCGGGACTATCCCCTGTGGATGTATTTGTTAGTACAGTGGA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 TGAGAAAGAAGGGAAACCGTCGGGACTATCCCCTGTGGATGTATTTGTTAGTACAGTGGA 1080

Qy       1081 TCCATTGAAAGAGCCTCCGCTTATTACTGCAAATACTGTCTTGTCTATTCTTGCTGTTGA 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 TCCATTGAAAGAGCCTCCGCTTATTACTGCAAATACTGTCTTGTCTATTCTTGCTGTTGA 1140

Qy       1141 TTATCCTGTCGATAAGGTTGCTTGTTACGTATCTGATGATGGTGCTGCTATGCTTACTTT 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 TTATCCTGTCGATAAGGTTGCTTGTTACGTATCTGATGATGGTGCTGCTATGCTTACTTT 1200

Qy       1201 CGAAGCTCTTTCTGAGACCGCTGAATTCGCAAGGAAATGGGTTCCTTTCTGCAAGAAATA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 CGAAGCTCTTTCTGAGACCGCTGAATTCGCAAGGAAATGGGTTCCTTTCTGCAAGAAATA 1260

Qy       1261 TTGTATTGAGCCTCGTGCTCCCGAATGGTATTTCTGCCATAAAATGGACTACTTGAAGAA 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 TTGTATTGAGCCTCGTGCTCCCGAATGGTATTTCTGCCATAAAATGGACTACTTGAAGAA 1320

Qy       1321 TAAAGTTCATCCCGCATTTGTTAGGGAGCGGCGAGCCATGAAGAGAGATTATGAAGAATT 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TAAAGTTCATCCCGCATTTGTTAGGGAGCGGCGAGCCATGAAGAGAGATTATGAAGAATT 1380

Qy       1381 CAAAGTAAAGATCAATGCTTTAGTAGCAACAGCACAGAAAGTGCCTGAGGATGGTTGGAC 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CAAAGTAAAGATCAATGCTTTAGTAGCAACAGCACAGAAAGTGCCTGAGGATGGTTGGAC 1440

Qy       1441 TATGCAAGACGGTACACCTTGGCCCGGTAATAGTGTGCGAGATCATCCTGGCATGATTCA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 TATGCAAGACGGTACACCTTGGCCCGGTAATAGTGTGCGAGATCATCCTGGCATGATTCA 1500

Qy       1501 GGTCTTCCTTGGAAGTGACGGTGTTCGTGATGTCGAAAACAACGAGTTGCCTCGATTAGT 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 GGTCTTCCTTGGAAGTGACGGTGTTCGTGATGTCGAAAACAACGAGTTGCCTCGATTAGT 1560

Qy       1561 TTACGTTTCTCGTGAGAAGAGACCCGGATTTGATCACCATAAGAAGGCTGGAGCTATGAA 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 TTACGTTTCTCGTGAGAAGAGACCCGGATTTGATCACCATAAGAAGGCTGGAGCTATGAA 1620

Qy       1621 TTCCCTGATACGAGTCTCTGGGGTTCTATCAAATGCTCCTTACCTTCTGAATGTCGATTG 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TTCCCTGATACGAGTCTCTGGGGTTCTATCAAATGCTCCTTACCTTCTGAATGTCGATTG 1680
```

```
Qy       1681 TGATCACTACATCAACAATAGCAAAGCTCTTAGAGAAGCAATGTGTTTCATGATGGATCC 1740
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TGATCACTACATCAACAATAGCAAAGCTCTTAGAGAAGCAATGTGTTTCATGATGGATCC 1740

Qy       1741 TCAGTCAGGAAAGAAAATCTGTTATGTTCAGTTCCCTCAAAGGTTCGATGGGATTGATAG 1800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 TCAGTCAGGAAAGAAAATCTGTTATGTTCAGTTCCCTCAAAGGTTCGATGGGATTGATAG 1800

Qy       1801 GCACGATCGATACTCAAATCGCAATGTTGTGTTCTTTGATATCAATATGAAAGGTTTGGA 1860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 GCACGATCGATACTCAAATCGCAATGTTGTGTTCTTTGATATCAATATGAAAGGTTTGGA 1860

Qy       1861 TGGGCTACAAGGGCCTATATACGTCGGTACAGGTTGTGTTTTCAGGAGGCAAGCGCTTTA 1920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 TGGGCTACAAGGGCCTATATACGTCGGTACAGGTTGTGTTTTCAGGAGGCAAGCGCTTTA 1920

Qy       1921 CGGATTTGATGCACCGAAGAAGAAGAAGGGCCCACGTAAGACATGCAATTGCTGGCCAAA 1980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 CGGATTTGATGCACCGAAGAAGAAGAAGGGCCCACGTAAGACATGCAATTGCTGGCCAAA 1980

Qy       1981 ATGGTGTCTCCTATGTTTTGGTTCAAGAAAGAATCGTAAAGCAAAGACAGTGGCTGCGGA 2040
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 ATGGTGTCTCCTATGTTTTGGTTCAAGAAAGAATCGTAAAGCAAAGACAGTGGCTGCGGA 2040

Qy       2041 TAAGAAGAAGAAGAATAGGGAAGCGTCAAAGCAGATCCACGCATTAGAAAATATCGAAGA 2100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 TAAGAAGAAGAAGAATAGGGAAGCGTCAAAGCAGATCCACGCATTAGAAAATATCGAAGA 2100

Qy       2101 GGGCCGCGGTCATAAAGTTCTTAACGTAGAACAGTCAACCGAGGCAATGCAAATGAAGTT 2160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 GGGCCGCGGTCATAAAGTTCTTAACGTAGAACAGTCAACCGAGGCAATGCAAATGAAGTT 2160

Qy       2161 GCAGAAGAAATATGGGCAGTCTCCTGTATTTGTTGCATCTGCGCGTCTGGAGAATGGTGG 2220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 GCAGAAGAAATATGGGCAGTCTCCTGTATTTGTTGCATCTGCGCGTCTGGAGAATGGTGG 2220

Qy       2221 GATGGCTAGAAACGCAAGCCCGGCTTGTCTGCTTAAAGAAGCCATCCAAGTCATTAGTCG 2280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 GATGGCTAGAAACGCAAGCCCGGCTTGTCTGCTTAAAGAAGCCATCCAAGTCATTAGTCG 2280

Qy       2281 CGGATATGAAGATAAAACTGAATGGGGAAAAGAGATTGGGTGGATCTATGGTTCTGTTAC 2340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 CGGATATGAAGATAAAACTGAATGGGGAAAAGAGATTGGGTGGATCTATGGTTCTGTTAC 2340

Qy       2341 CGAAGATATTCTTACGGGTTCTAAGATGCATTCTCATGGTTGGAGACATGTTTATTGTAC 2400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 CGAAGATATTCTTACGGGTTCTAAGATGCATTCTCATGGTTGGAGACATGTTTATTGTAC 2400

Qy       2401 ACCAAAGTTAGCGGCTTTCAAAGGATCAGCTCCAATCAATCTTCGGATCGTCTCCATCA 2460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 ACCAAAGTTAGCGGCTTTCAAAGGATCAGCTCCAATCAATCTTCGGATCGTCTCCATCA 2460

Qy       2461 AGTTCTTCGATGGGCGCTTGGGTCGGTTGAGATTTTCTTGAGTAGGCATTGTCCTATTTG 2520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2461 AGTTCTTCGATGGGCGCTTGGGTCGGTTGAGATTTTCTTGAGTAGGCATTGTCCTATTTG 2520

Qy       2521 GTATGGTTATGGAGGTGGGTTGAAATGGCTTGAGCGGTTGTCCTACATTAACTCTGTGGT 2580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2521 GTATGGTTATGGAGGTGGGTTGAAATGGCTTGAGCGGTTGTCCTACATTAACTCTGTGGT 2580

Qy       2581 TTACCCGTGGACCTCTCTACCGCTCATCGTTTACTGTTCTCTCCCTGCCATCTGTCTTCT 2640
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2581 TTACCCGTGGACCTCTCTACCGCTCATCGTTTACTGTTCTCTCCCTGCCATCTGTCTTCT 2640

Qy       2641 CACTGGAAAATTCATCGTTCCCGAGATTAGCAACTATGCGAGTATCCTCTTCATGGCGCT 2700
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2641 CACTGGAAAATTCATCGTTCCCGAGATTAGCAACTATGCGAGTATCCTCTTCATGGCGCT 2700

Qy       2701 CTTCTCGTCGATTGCAATAACGGGTATTCTCGAGATGCAATGGGGCAAAGTTGGGATCGA 2760
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2701 CTTCTCGTCGATTGCAATAACGGGTATTCTCGAGATGCAATGGGGCAAAGTTGGGATCGA 2760

Qy       2761 TGATTGGTGGAGAAACGAACAGTTTTGGGTCATTGGAGGTGTTTCTGCGCATCTGTTTGC 2820
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2761 TGATTGGTGGAGAAACGAACAGTTTTGGGTCATTGGAGGTGTTTCTGCGCATCTGTTTGC 2820

Qy       2821 TCTCTTCCAAGGTCTCCTCAAGGTTCTTGCTGGTGTCGACACTAACTTCACAGTCACATC 2880
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2821 TCTCTTCCAAGGTCTCCTCAAGGTTCTTGCTGGTGTCGACACTAACTTCACAGTCACATC 2880

Qy       2881 AAAAGCAGCTGATGATGGAGAGTTCTCTGACCTTTACCTCTTCAAATGGACTTCACTTCT 2940
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2881 AAAAGCAGCTGATGATGGAGAGTTCTCTGACCTTTACCTCTTCAAATGGACTTCACTTCT 2940

Qy       2941 CATCCCTCCAATGACTCTACTCATCATAAACGTCATTGGAGTCATAGTCGGAGTCTCTGA 3000
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2941 CATCCCTCCAATGACTCTACTCATCATAAACGTCATTGGAGTCATAGTCGGAGTCTCTGA 3000

Qy       3001 TGCCATCAGCAATGGATACGACTCGTGGGGACCGCTTTTCGGAAGACTGTTCTTTGCACT 3060
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3001 TGCCATCAGCAATGGATACGACTCGTGGGGACCGCTTTTCGGAAGACTGTTCTTTGCACT 3060

Qy       3061 TTGGGTCATCATTCATCTTTACCCGTTCCTTAAAGGTTTGCTTGGGAAACAAGATAGAAT 3120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3061 TTGGGTCATCATTCATCTTTACCCGTTCCTTAAAGGTTTGCTTGGGAAACAAGATAGAAT 3120

Qy       3121 GCCAACCATTATTGTCGTCTGGTCCATCCTCCTGGCCTCGATTCTTACACTTCTTTGGGT 3180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3121 GCCAACCATTATTGTCGTCTGGTCCATCCTCCTGGCCTCGATTCTTACACTTCTTTGGGT 3180

Qy       3181 CCGGGTTAATCCGTTTGTGGCGAAAGGCGGTCCTATTCTCGAGATCTGTGGTTTAGACTG 3240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3181 CCGGGTTAATCCGTTTGTGGCGAAAGGCGGTCCTATTCTCGAGATCTGTGGTTTAGACTG 3240

Qy       3241 CTTGTGATTCGATTGACCGGTGGATGGGTTGGTGAAAAAGGTTTAATTCCCACGGATCAA 3300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3241 CTTGTGATTCGATTGACCGGTGGATGGGTTGGTGAAAAAGGTTTAATTCCCACGGATCAA 3300

Qy       3301 AGAGAGGTAAGAGAGATATTGTTTTACCTCTAAAAGACTCCTTCATTGTGTTCATTAGAT 3360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3301 AGAGAGGTAAGAGAGATATTGTTTTACCTCTAAAAGACTCCTTCATTGTGTTCATTAGAT 3360

Qy       3361 GATGAAAAATGAAAAGAAAAAGAAGATTTAATTTTGTTACGAGAATTGTTATTTTTGCAA 3420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3361 GATGAAAAATGAAAAGAAAAAGAAGATTTAATTTTGTTACGAGAATTGTTATTTTTGCAA 3420

Qy       3421 GAATGTGTTGTAGATAGCGGCCGC 3444
             ||||||||||||||||||||||||
Db       3421 GAATGTGTTGTAGATAGCGGCCGC 3444
```

RESULT 2
ABZ14725
ID ABZ14725 standard; DNA; 3255 BP.
XX
AC ABZ14725;
XX
DT 21-JAN-2003 (first entry)
XX
DE Arabidopsis thaliana stress regulated gene SEQ ID NO 2530.
XX
KW Arabidopsis thaliana; plant; gene; stress; transgenic; ds.
XX
OS Arabidopsis thaliana.
XX
PN WO200216655-A2.
XX
PD 28-FEB-2002.
XX
PF 24-AUG-2001; 2001WO-US26685.
XX
PR 24-AUG-2000; 2000US-227866P.
PR 26-JAN-2001; 2001US-264647P.
PR 22-JUN-2001; 2001US-300111P.
XX

```
A   (SCRI ) SCRIPPS RES INST.
A   (SYGN ) SYNGENTA PARTICIPATIONS AG.
X
I   Harper JF, Kreps J, Wang X, Zhu T;
X
R   WPI; 2002-304127/34.
X
T   Identifying a stress condition to which a plant cell has been exposed
T   and producing plants with increased tolerance to these abiotic stresses
T   -
X
S   Claim 144; SEQ ID NO 2530; 577pp + Sequence Listing; English.
X
C   The invention relates to identifying a stress condition to which a plant
C   cell has been exposed, comprising:
C   (a) contacting nucleic acid representative of expressed polynucleotides
C   in the plant cell with an array or probes representative of the plant
C   cell genome; and
C   (b) detecting a profile of expressed polynucleotides in the plant cell
C   characteristic of a stress response. The method is useful in the
C   production of transgenic plants, cells and seeds and in producing plants
C   with increased tolerance to abiotic stress. The present sequence is that
C   of an Arabidopsis thaliana stress regulated gene (ABZ12196-ABZ17574) used
C   in methods of the invention.
C   Note: The sequence data for this patent is not represented in the printed
C   specification but is based on sequence information supplied to Derwent by
C   the European Patent Office.
X
Q   Sequence 3255 BP; 841 A; 641 C; 813 G; 960 T; 0 other;

 Query Match            93.2%;  Score 3211.2;  DB 24;  Length 3255;
 Best Local Similarity  99.4%;  Pred. No. 0;
 Matches 3222;  Conservative    0;  Mismatches   18;  Indels    0;  Gaps    0;

Qy          8 CGGTTAATCGCCGGTTCTCACAACAGGAATGAGTTTGTCCTCATTAATGCCGATGAGAAT 67
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         16 CGGTTAATCGCCGGTTCTCACAACAGGAATGAGTTTGTCCTCATTAATGCCGATGAGAAT 75

Qy         68 GCCCGAATAAGATCAGTCCAAGAGCTGAGTGGACAGACATGTCAAATCTGCAGAGATGAG 127
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         76 GCCCGAATAAGATCAGTCCAAGAGCTGAGTGGACAGACATGTCAAATCTGCAGAGATGAG 135

Qy        128 ATCGAATTGACTGTTGATGGAGAACCGTTTGTGGCATGTAACGAATGTGCATTCCCTGTG 187
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        136 ATCGAATTGACTGTTGATGGAGAACCGTTTGTGGCATGTAACGAATGTGCATTCCCTGTG 195

Qy        188 TGTAGACCTTGCTATGAGTACGAAAGACGAGAAGGCAATCAAGCTTGTCCACAGTGCAAA 247
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        196 TGTAGACCTTGCTATGAGTACGAAAGACGAGAAGGCAATCAAGCTTGTCCACAGTGCAAA 255

Qy        248 ACCCGTTTCAAACGTCTTAAAGGAAGTCCAAGAGTTGAAGGTGATGAAGAGGAAGATGAC 307
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        256 ACCCGTTTCAAACGTCTTAAAGGAAGTCCAAGAGTTGAAGGTGATGAAGAGGAAGATGAC 315

Qy        308 ATTGATGATTTAGACAATGAGTTTGAGTATGGAAATAATGGGATTGGATTTGATCAGGTT 367
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        316 ATTGATGATTTAGACAATGAGTTTGAGTATGGAAATAATGGGATTGGATTTGATCAGGTT 375

Qy        368 TCTGAAGGTATGTCAATCTCTCGTCGCAACTCCGGTTTCCCACAATCTGATTTGGATTCA 427
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        376 TCTGAAGGTATGTCAATCTCTCGTCGCAACTCCGGTTTCCCACAATCTGATTTGGATTCA 435

Qy        428 GCTCCACCTGGCTCTCAGATTCCATTGCTGACTTACGGCGACGAGGACGTTGAGATTTCT 487
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        436 GCTCCACCTGGCTCTCAGATTCCATTGCTGACTTACGGCGACGAGGACGTTGAGATTTCT 495

Qy        488 TCTGATAGACATGCTCTTATTGTTCCTCCTTCACTTGGTGGTCATGGCAATAGAGTTCAT 547
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        496 TCTGATAGACATGCTCTTATTGTTCCTCCTTCACTTGGTGGTCATGGCAATAGAGTTCAT 555

Qy        548 CCTGTTTCTCTTTCTGACCCGACCGTGGCTGCACATCGAAGGCTGATGGTACCTCAGAAA 607
              |||||||||||||||||||||||||||||||| |||||  ||||||||||||||||||
Db        556 CCTGTTTCTCTTTCTGACCCGACCGTGGCTGCACATCCAAGGCCTATGGTACCTCAGAAA 615

Qy        608 GATCTTGCGGTTTATGGTTATGGAAGTGTCGCTTGGAAAGATCGGATGGAGGAATGGAAG 667
              ||||||||||||:|||||||||||||||||||:|||||||||||||||||||||||||||
Db        616 GATCTTGCGGTTTATGGTTATGGAAGTGTCGCTTGGAAAGATCGGATGGAGGAATGGAAG 675

Qy        668 AGAAAGCAGAATGAGAAACTTCAGGTTGTTAGGCATGAAGGAGATCCTGATTTTGAAGAT 727
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        676 AGAAAGCAGAATGAGAAACTTCAGGTTGTTAGGCATGAAGGAGATCCTGATTTTGAAGAT 735

Qy        728 GGTGATGATGCTGATTTTCCAATGATGGATGAGGGAAGGCAGCCATTGTCTATGAAGATA 787
              |||||||||||||:||||:|||||||||||||||||||||||||||||||||| ||||||
Db        736 GGTGATGATGCTGATTTTCCAATGATGGATGAGGGAAGGCAGCCATTGTCTAGGAAGATA 795

Qy        788 CCAATCAAATCGAGCAAGATAAATCCTTACCGGATGTTAATTGTGCTACGTCTTGTGATT 847
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        796 CCAATCAAATCGAGCAAGATAAATCCTTACCGGATGTTAATTGTGCTACGTCTTGTGATT 855

Qy        848 CTTGGTCTCTTCTTTCACTACCGTATTCTTCACCCCGTCAAAGATGCATATGCTTTGTGG 907
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        856 CTTGGTCTCTTCTTTCACTACCGTATTCTTCACCCCGTCAAAGATGCATATGCTTTGTGG 915

Qy        908 CTTATTTCTGTTATATGTGAGATATGGTTTGCTGTTTCATGGGTTCTTGATCAGTTCCCT 967
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        916 CTTATTTCTGTTATATGTGAGATATGGTTTGCTGTTTCATGGGTTCTTGATCAGTTCCCT 975

Qy        968 AAATGGTACCCTATCGAGCGAGAAACGTACTTGGACCGACTCTCATTAAGATATGAGAAA 1027
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        976 AAATGGTACCCTATCGAGCGAGAAACGTACTTGGACCGACTCTCATTAAGATATGAGAAA 1035

Qy       1028 GAAGGGAAACCGTCGGGACTATCCCCTGTGGATGTATTTGTTAGTACAGTGGATCCATTG 1087
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1036 GAAGGGAAACCGTCGGGACTATCCCCTGTGGATGTATTTGTTAGTACAGTGGATCCATTG 1095

Qy       1088 AAAGAGCCTCCGCTTATTACTGCAAATACTGTCTTGTCTATTCTTGCTGTTGATTATCCT 1147
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1096 AAAGAGCCTCCGCTTATTACTGCAAATACTGTCTTGTCTATTCTTGCTGTTGATTATCCT 1155

Qy       1148 GTCGATAAGGTTGCTTGTTACGTATCTGATGATGGTGCTGCTATGCTTACTTTCGAAGCT 1207
              |||||||||||||:|||||||||||||||||||||||||||||||||||||||||||||||
Db       1156 GTCGATAAGGTTGCTTGTTACGTATCTGATGATGGTGCTGCTATGCTTACTTTCGAAGCT 1215

Qy       1208 CTTTCTGAGACCGCTGAATTCGCAAGGAAATGGGTTCCTTTCTGCAAGAAATATTGTATT 1267
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1216 CTTTCTGAGACCGCTGAATTCGCAAGGAAATGGGTTCCTTTCTGCAAGAAATATTGTATT 1275

Qy       1268 GAGCCTCGTGCTCCCGAATGGTATTTCTGCCATAAAATGGACTACTTGAAGAATAAAGTT 1327
              |||||||||||||:|||||||||||||||||||||||||||||||||||||||||||||||
Db       1276 GAGCCTCGTGCTCCCGAATGGTATTTCTGCCATAAAATGGACTACTTGAAGAATAAAGTT 1335

Qy       1328 CATCCCGCATTTGTTAGGGAGCGGCGAGCCATGAAGAGAGATTATGAAGAATTCAAAGTA 1387
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1336 CATCCCGCATTTGTTAGGGAGCGGCGAGCCATGAAGAGAGATTATGAAGAATTCAAAGTA 1395

Qy       1388 AAGATCAATGCTTTAGTAGCAACAGCACAGAAAGTGCCTGAGGATGGTTGGACTATGCAA 1447
              |||||||||||||:|||||||||||||||||||||||||||||||||||||||||||||||
Db       1396 AAGATCAATGCTTTAGTAGCAACAGCACAGAAAGTGCCTGAGGATGGTTGGACTATGCAA 1455

Qy       1448 GACGGTACACCTTGGCCCGGTAATAGTGTGCGAGATCATCCTGGCATGATTCAGGTCTTC 1507
              |||||||||||||:|||||||||||||||||||||||||||||||||||||||||||||||
Db       1456 GACGGTACACCTTGGCCCGGTAATAGTGTGCGAGATCATCCTGGCATGATTCAGGTCTTC 1515

Qy       1508 CTTGGAAGTGACGGTGTTCGTGATGTCGAAAACAACGAGTTGCCTCGATTAGTTTACGTT 1567
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1516 CTTGGAAGTGACGGTGTTCGTGATGTCGAAAACAACGAGTTGCCTCGATTAGTTTACGTT 1575

Qy       1568 TCTCGTGAGAAGAGACCCGGATTTGATCACCATAAGAAGGCTGGAGCTATGAATTCCCTG 1627
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1576 TCTCGTGAGAAGAGACCCGGATTTGATCACCATAAGAAGGCTGGAGCTATGAATTCCCTG 1635

Qy       1628 ATACGAGTCTCTGGGGTTCTATCAAATGCTCCTTACCTTCTGAATGTCGATTGTGATCAC 1687
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1636 ATACGAGTCTCTGGGGTTCTATCAAATGCTCCTTACCTTCTGAATGTCGATTGTGATCAC 1695
```

```
Qy       1688 TACATCAACAATAGCAAAGCTCTTAGAGAAGCAATGTGTTTCATGATGGATCCTCAGTCA 1747
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1696 TACATCAACAATAGCAAAGCTCTTAGAGAAGCAATGTGTTTCATGATGGATCCTCAGTCA 1755

Qy       1748 GGAAAGAAAAATCTGTTATGTTCAGTTCCCTCAAAGGTTCGATGGGATTGATAGGCACGAT 1807
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1756 GGAAAGAAAAATCTGTTATGTTCAGTTCCCTCAAAGGTTCGATGGGATTGATAGGCACGAT 1815

Qy       1808 CGATACTCAAATCGCAATGTTGTGTTCTTTGATATCAATATGAAAGGTTTGGATGGGCTA 1867
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1816 CGATACTCAAATCGCAATGTTGTGTTCTTTGATATCAATATGAAAGGTTTGGATGGGCTA 1875

Qy       1868 CAAGGGCCTATATACGTCGGTACAGGTTGTGTTTTCAGGAGGCAAGCGCTTTACGGATTT 1927
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1876 CAAGGGCCTATATACGTCGGTACAGGTTGTGTTTTCAGGAGGCAAGCGCTTTACGGATTT 1935

Qy       1928 GATGCACCGAAGAAGAAGAAGGGCCCACGTAAGACATGCAATTGCTGGCCAAAATGGTGT 1987
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1936 GATGCACCGAAGAAGAAGAAGGGCCCACGTAAGACATGCAATTGCTGGCCAAAATGGTGT 1995

Qy       1988 CTCCTATGTTTTGGTTCAAGAAAGAATCGTAAAGCAAAGACAGTGGCTGCGGATAAGAAG 2047
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1996 CTCCTATGTTTTGGTTCAAGAAAGAATCGTAAAGCAAAGACAGTGGCTGCGGATAAGAAG 2055

Qy       2048 AAGAAGAATAGGGAAGCGTCAAAGCAGATCCACGCATTAGAAAATATCGAAGAGGGCCGC 2107
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2056 AAGAAGAATAGGGAAGCGTCAAAGCAGATCCACGCATTAGAAAATATCGAAGAGGGCCGC 2115

Qy       2108 GGTCATAAAGTTCTTAACGTAGAACAGTCAACCGAGGCAATGCAAATGAAGTTGCAGAAG 2167
             |    |||||  |  ||||||||||||||||||||||||||||||||||||| |||||
Db       2116 GTCACTAAAGGTTCTAACGTAGAACAGTCAACCGAGGCAATGCAAATGAAGTTGGAGAAG 2175

Qy       2168 AAATATGGGCAGTCTCCTGTATTTGTTGCATCTGCGCGTCTGGAGAATGGTGGGATGGCT 2227
             ||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2176 AAATTTGGGCAGTCTCCTGTATTTGTTGCATCTGCGCGTATGGAGAATGGTGGGATGGCT 2235

Qy       2228 AGAAACGCAAGCCCGGCTTGTCTGCTTAAAGAAGCCATCCAAGTCATTAGTCGCGGATAT 2287
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2236 AGAAACGCAAGCCCGGCTTGTCTGCTTAAAGAAGCCATCCAAGTCATTAGTTGCGGATAT 2295

Qy       2288 GAAGATAAAACTGAATGGGGAAAAGAGATTGGGTGGATCTATGGTTCTGTTACCGAAGAT 2347
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2296 GAAGATAAAACTGAATGGGGAAAAGAGATTGGGTGGATCTATGGTTCTGTTACCGAAGAT 2355

Qy       2348 ATTCTTACGGGTTCTAAGATGCATTCTCATGGTTGGAGACATGTTTATTGTACACCAAAG 2407
             |||||||||||| |||||||||||||||||||||||  ||||||||||||||||||||||
Db       2356 ATTCTTACGGGTTTTAAGATGCATTCTCATGGTTGGAGATCTGTTTATTGTACACCAAAG 2415

Qy       2408 TTAGCGGCTTTCAAAGGATCAGGTCCAATCAATCTTTCGGATGGTCTCCATCAAGTTCTT 2467
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2416 TTAGCGGCTTTCAAAGGATCAGGTCCAATCAATCTTTCGGATGGTCTCCATCAAGTTCTT 2475

Qy       2468 CGATGGGCGCTTGGGTCGGTTGAGATTTTCTTGAGTAGGCATTGTCCTATTTGGTATGGT 2527
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2476 CGATGGGCGCTTGGGTCGGTTGAGATTTTCTTGAGTAGGCATTGTCCTATTTGGTATGGT 2535

Qy       2528 TATGGAGGTGGGTTGAAATGGCTTGAGCGGTTGTCCTACATTAACTCTGTGGTTTACCCG 2587
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2536 TATGGAGGTGGGTTGAAATGGCTTGAGCGGTTGTCCTACATTAACTCTGTGGTTTACCCG 2595

Qy       2588 TGGACCTCTCTACCGCTCATCGTTTACTGTTCTCTCCCTGCCATCTGTCTTCTCACTGGA 2647
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2596 TGGACCTCTCTACCGCTCATCGTTTACTGTTCTCTCCCTGCCATCTGTCTTCTCACTGGA 2655

Qy       2648 AAATTCATCGTTCCCGAGATTAGCAACTATGCGAGTATCCTCTTCATGGCGCTCTTCTCG 2707
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2656 AAATTCATCGTTCCCGAGATTAGCAACTATGCGAGTATCCTCTTCATGGCGCTCTTCTCG 2715

Qy       2708 TCGATTGCAATAACGGGTATTCTCGAGATGCAATGGGGCAAAGTTGGGATCGATGATTGG 2767
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2716 TCGATTGCAATAACGGGTATTCTCGAGATGCAATGGGGCAAAGTTGGGATCGATGATTGG 2775

Qy       2768 TGGAGAAACGAACAGTTTTGGGTCATTGGAGGTGTTTCTGCGCATCTGTTTGCTCTCTTC 2827
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2776 TGGAGAAACGAACAGTTTTGGGTCATTGGAGGTGTTTCTGCGCATCTGTTTGCTCTCTTC 2835

Qy       2828 CAAGGTCTCCTCAAGGTTCTTGCTGGTGTCGACACTAACTTCACAGTCACATCAAAAGCA 2887
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2836 CAAGGTCTCCTCAAGGTTCTTGCTGGTGTCGACACTAACTTCACAGTCACATCAAAAGCA 2895

Qy       2888 GCTGATGATGGAGAGTTCTCTGACCTTTACCTCTTCAAATGGACTTCACTTCTCATCCCT 2947
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2896 GCTGATGATGGAGAGTTCTCTGACCTTTACCTCTTCAAATGGACTTCACTTCTCATCCCT 2955

Qy       2948 CCAATGACTCTACTCATCATAAACGTCATTGGAGTCATAGTCGGAGTCTCTGATGCCATC 3007
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2956 CCAATGACTCTACTCATCATAAACGTCATTGGAGTCATAGTCGGAGTCTCTGATGCCATC 3015

Qy       3008 AGCAATGGATACGACTCGTGGGGACCGCTTTTCGGAAGACTGTTCTTTGCACTTTGGGTC 3067
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3016 AGCAATGGATACGACTCGTGGGGACCGCTTTTCGGAAGACTGTTCTTTGCACTTTGGGTC 3075

Qy       3068 ATCATTCATCTTTACCCGTTCCTTAAAGGTTTGCTTGGGAAACAAGATAGAATGCCAACC 3127
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3076 ATCATTCATCTTTACCCGTTCCTTAAAGGTTTGCTTGGGAAACAAGATAGAATGCCAACC 3135

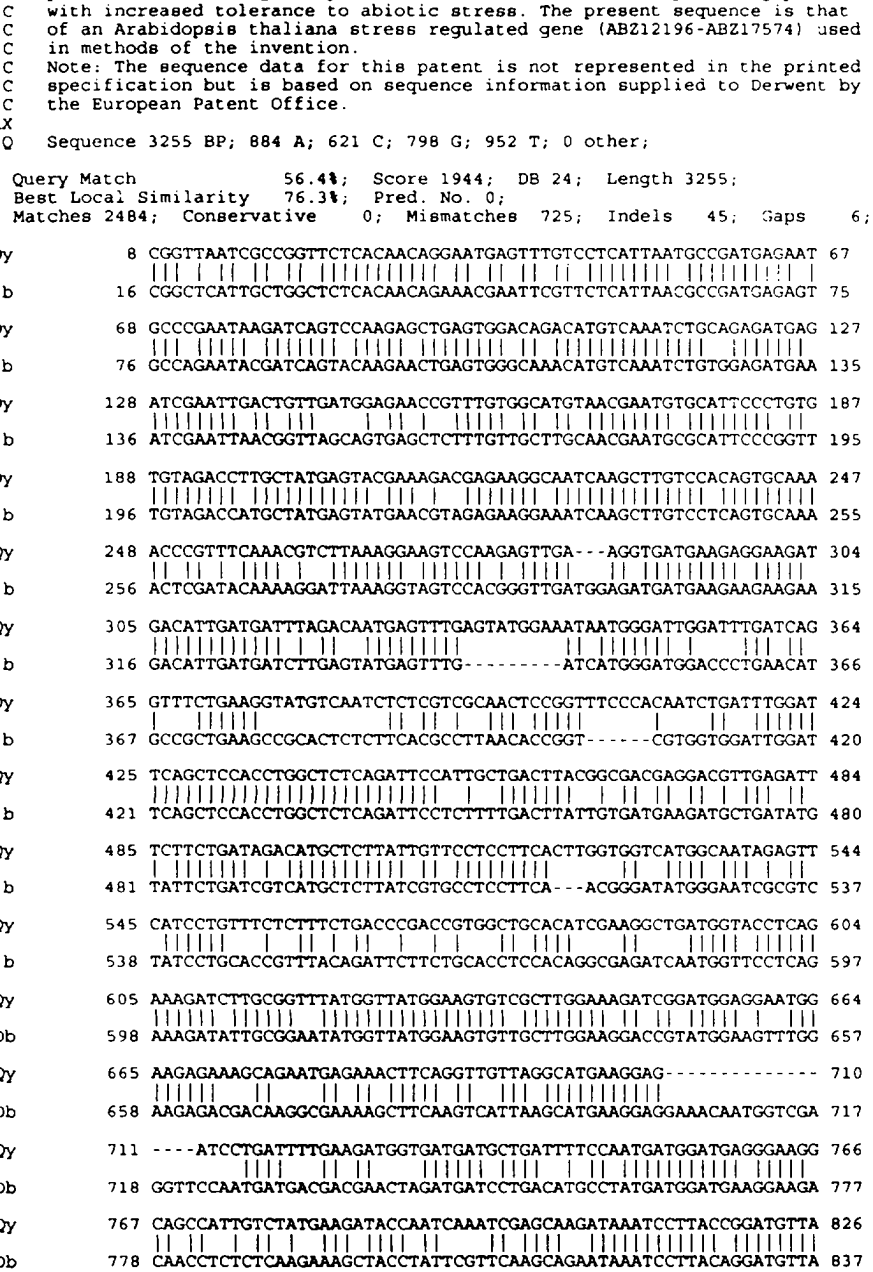
Qy       3128 ATTATTGTCGTCTGGTCCATCCTCCTGGCCTCGATTCTTACACTTCTTTGGGTCCGGGTT 3187
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3136 ATTATTGTCGTCTGGTCCATCCTCCTGGCCTCGATTCTTACACTTCTTTGGGTCCGGGTT 3195

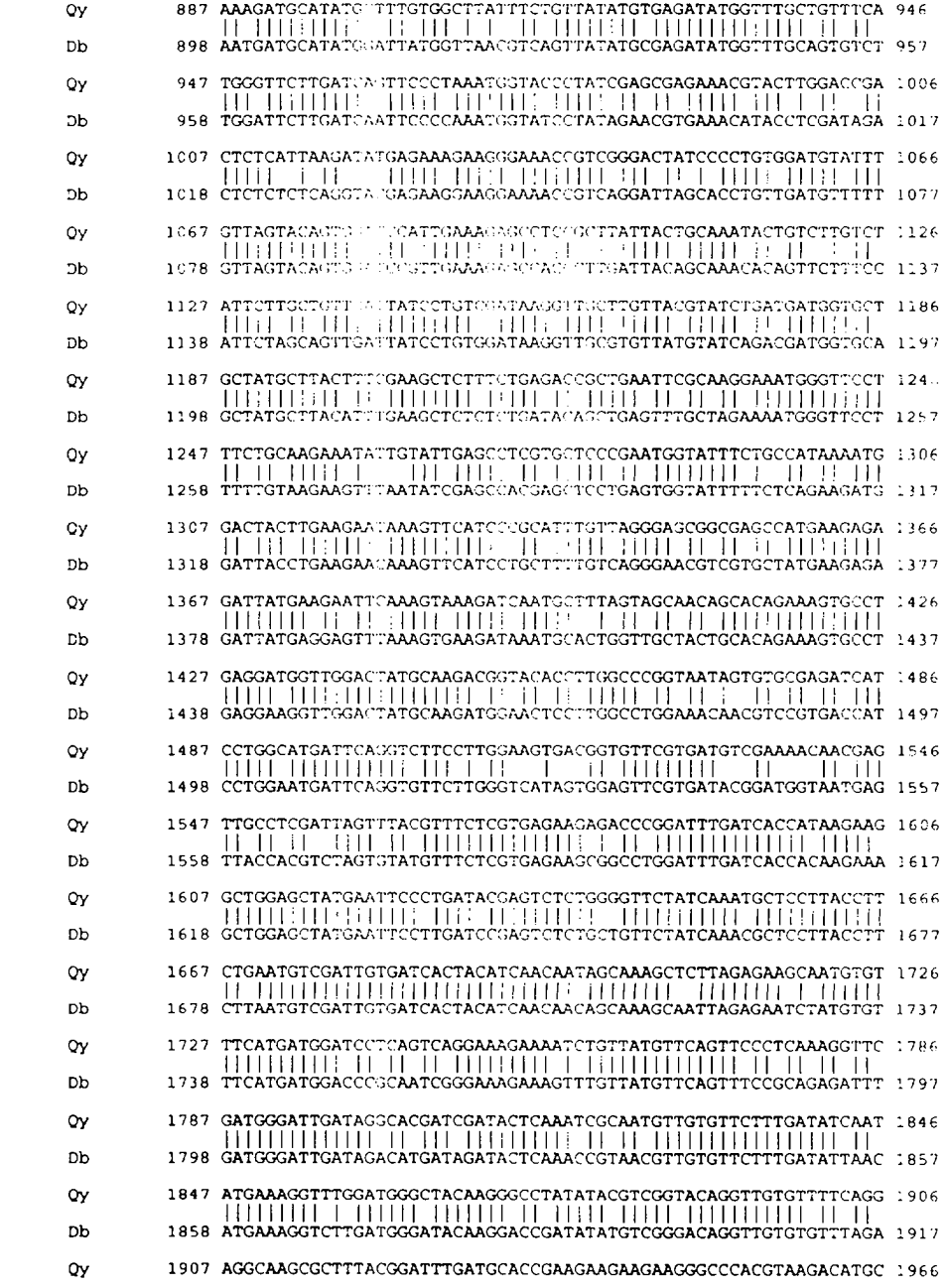
Qy       3188 AATCCGTTTGTGGCGAAAGGCGGTCCTATTCTCGAGATCTGTGGTTTAGACTGCTTGTGA 3247
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3196 AATCCGTTTGTGGCGAAAGGCGGTCCTATTCTCGAGATCTGTGGTTTAGACTGCTTGTGA 3255
```

RESULT 3
ABZ12754
ID ABZ12754 standard; DNA; 3255 BP.
XX
AC ABZ12754;
XX
DT 21-JAN-2003 (first entry)
XX
DE Arabidopsis thaliana stress regulated gene SEQ ID NO 559.
XX
KW Arabidopsis thaliana; plant; gene; stress; transgenic; ds.
XX
OS Arabidopsis thaliana.
XX
PN WO200216655-A2.
XX
PD 28-FEB-2002.
XX
PF 24-AUG-2001; 2001WO-US26685.
XX
PR 24-AUG-2000; 2000US-227866P.
PR 26-JAN-2001; 2001US-264647P.
PR 22-JUN-2001; 2001US-300111P.
XX
PA (SCRI) SCRIPPS RES INST.
PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
PI Harper JF, Kreps J, Wang X, Zhu T;
XX
DR WPI; 2002-304127/34.
XX
PT Identifying a stress condition to which a plant cell has been exposed
PT and producing plants with increased tolerance to these abiotic stresses
PT -
XX
PS Claim 144; SEQ ID NO 559; 577pp + Sequence Listing; English.
XX
CC The invention relates to identifying a stress condition to which a plant
CC cell has been exposed, comprising:
CC (a) contacting nucleic acid representative of expressed polynucleotides
CC in the plant cell with an array or probes representative of the plant

C cell genome; and
C (b) detecting a profile of expressed polynucleotides in the plant cell
C characteristic of a stress response. The method is useful in the
C production of transgenic plants, cells and seeds and in producing plants
C with increased tolerance to abiotic stress. The present sequence is that
C of an Arabidopsis thaliana stress regulated gene (ABZ12196-ABZ17574) used
C in methods of the invention.
C Note: The sequence data for this patent is not represented in the printed
C specification but is based on sequence information supplied to Derwent by
C the European Patent Office.
X
Q Sequence 3255 BP; 884 A; 621 C; 798 G; 952 T; 0 other;

Query Match 56.4%; Score 1944; DB 24; Length 3255;
Best Local Similarity 76.3%; Pred. No. 0;
Matches 2484; Conservative 0; Mismatches 725; Indels 45; Gaps 6;





```
Db       1918 AGACAGGCTCTTTATGGTTTTGATGCACCAAAGAAGAAGAAACCACCAGGCAAAACCTGT 1977

Qy       1967 AATTGCTGGCCAAAATGGTGTCTCCTATGTTTTGGTTCAAGAAAGAATCGTAAAGCAAAG 2026

Db       1978 AACTGTTGGCCTAAATGGTGTTGTTTGTGTTGTGGGTTGAGAAAGAAGAGTAAAACGAAA 2037

Qy       2027 ACAGTGGCTGCGGATAAGAAGAAGAAGAATAGGGAAGCGTCAAAGCAGATCCACGCATTA 2086

Db       2038 ------GCCAAAGATAAGAAAACTAACACTAAAGAGACTTCAAAGCAGATTCATGCGCTA 2091

Qy       2087 GAAAATATCGAAGAGGGCCGCGGTCATAAAGTTCTTAACGTAGAACAGTCAACCGAGGCA 2146

Db       2092 GAGAATGTCGACGAAGGTGTTATCGTCCCAGTGTCAAATGTTGAGAAGAGATCTGAAGCA 2151

Qy       2147 ATGCAAATGAAGTTGCAGAAGAAATATGGGCAGTCTCCTGTATTTGTTGCATCTGCGCGT 2206

Db       2152 ACACAATTGAAATTGGAGAAGAAGTTTGGACAATCTCCGGTTTTCGTTGCCTCTGCTGTT 2211

Qy       2207 CTGGAGAATGGTGGGATGGCTAGAAACGCAAGCCCGGCTTGTCTGCTTAAAGAAGCCATC 2266

Db       2212 CTACAGAACGGTGGAGTTCCCCGTAACGCAAGCCCCGCATGTTTGTTAAGAGAAGCCATT 2271

Qy       2267 CAAGTCATTAGTCGCGGATATGAAGATAAAACTGAATGGGGAAAAAGAGATTGGGTGGATC 2326

Db       2272 CAAGTTATTAGCTGCGGGTACGAAGATAAAACCGAATGGGGAAAAGAGATCGGGTGGATT 2331

Qy       2327 TATGGTTCTGTTACCGAAGATATTCTTACGGGTTCTAAGATGCATTCTCATGGTTGGAGA 2386

Db       2332 TATGGATCGGTGACTGAAGATATCCTGACGGGTTTCAAGATGCATTGCCATGGATGGAGA 2391

Qy       2387 CATGTTTATTGTACACCAAAGTTAGCGGCTTTCAAAGGATCAGCTCCAATCAATCTTTCG 2446

Db       2392 TCTGTGTACTGTATGCCTAAGCGTGCAGCTTTTAAAGGATCTGCTCCTATTAACTTGTCA 2451

Qy       2447 GATCGTCTCCATCAAGTTCTTCGATGGGCGCTTGGGTCGGTTGAGATTTTCTTGAGTAGG 2506

Db       2452 GATCGTCTTCATCAAGTTCTACGTTGGGCTCTTGGCTCTGTAGAGATTTTCTTGAGCAGA 2511

Qy       2507 CATTGTCCTATTTGGTATGGTTATGGAGGTGGGTTGAAATGGCTTGAGCGGTTGTCCTAC 2566

Db       2512 CATTGTCCGATATGGTATGGTTATGGTGGTGGTTTAAAATGGTTGGAGAGATTCTCTTAC 2571

Qy       2567 ATTAACTCTGTGGTTTACCCGTGGACCTCTCTACCGCTCATCGTTTACTGTTCTCTCCCT 2626

Db       2572 ATCAACTCTGTCGTCTATCCTTGGACTTCACTTCCATTGATCGTCTATTGTTCTCTCCCC 2631

Qy       2627 GCCATCTGTCTTCTCACTGGAAAATTCATCGTTCCCGAGATTAGCAACTATGCGAGTATC 2686

Db       2632 GCGGTTTGTTTACTCACAGGAAAATTCATCGTCCCTGAGATAAGCAACTACGCAGGTATA 2691

Qy       2687 CTCTTCATGGCGCTCTTCTCGTCGATTGCAATAACGGGTATTTCGAGATGCAATGGGGC 2746

Db       2692 CTCTTCATGCTCATGTTCATATCCATAGCAGTAACTGGAATTCTCGAAATGCAATGGGGA 2751

Qy       2747 AAAGTTGGGATCGATGATTGGTGGAGAAACGAACAGTTTTGGGTCATTGGAGGTGTTTCT 2806

Db       2752 GGTGTCGGAATCGATGATTGGTGGAGAAACGAGCAGTTTTGGGTAATCGGAGGGGCCTCC 2811

Qy       2807 GCGCATCTGTTTGCTCTCTTCCAAGGTCTCCTCAAGGTTCTTGCTGGTGTCGACACTAAC 2866

Db       2812 TCGCATCTATTTGCTCTGTTTCAAGGTTGCTCAAAGTTCTAGCCGGAGTTAACACGAAT 2871

Qy       2867 TTCACAGTCACATCAAAAGCAGCTGATGATGGAGAGTTCTCTGACCTTTACCTCTTCAAA 2926

Db       2872 TTCACAGTCACTTCAAAAGCAGCAGACGATGGAGCTTTCTCTGAGCTTTACATCTTCAAG 2931

Qy       2927 TGGACTTCACTTCTCATCCCTCCAATGACTCTACTCATCATAAACGTCATTGGAGTCATA 2986

Db       2932 TGGACAACTTTGTTGATTCCTCCGACAACACTTCTGATCATTAACATCATTGGAGTTATT 2991

Qy       2987 GTCGGAGTCTCTGATGCCATCAGCAATGGATACGACTCGTGGGGACCGCTTTTCGGAAGA 3046

Db       2992 GTCGGCGTTTCTGATGCCATTAGCAATGGCTATGACTCATGGGGACCTCTCTTTGGGAGA 3051

Qy       3047 CTGTTCTTTGCACTTTGGGTCATCATTCATCTTTACCCGTTCCTTAAAGGTTTGCTTGGG 3106

Db       3052 CTTTTCTTCGCTCTTTGGGTCATTGTTCATTTATACCCATTCCTCAAGGGAATGCTTGGG 3111

Qy       3107 AAACAAGATAGAATGCCAACCATTATTGTCGTCTGGTCCATCCTCCTGGCCTCGATTCTT 3166

Db       3112 AAGCAAGACAAAATGCCTACGATTATTGTGGTCTGGTCTATTCTTCTAGCTTCGATCTTG 3171

Qy       3167 ACACTTCTTTGGGTCCGGGTTAATCCGTTTGTGGCGAAAGGCGGTCCTATTCTCGAGATC 3226

Db       3172 ACACTCTTGTGGGTCAGAGTTAACCCGTTTGTGGCTAAAGGGGGACCAGTGTTGGAGATC 3231

Qy       3227 TGTGGTTTAGACTG 3240

Db       3232 TGTGGTCTGAATTG 3245
```

RESULT 4
AAV06566
ID AAV06566 standard; cDNA; 3828 BP.
XX
AC AAV06566;
XX
DT 06-JUL-1998 (first entry)
XX
DE Arabidopsis cellulose biosynthetic gene clone Ath-A.
XX
KW Cellulose; cellulose synthase; RSW1 gene; beta-1,4-glucan;
KW transgenic plant; ss.
XX
OS Arabidopsis thaliana var. Columbia.
XX
FH Key Location/Qualifiers
FT CDS 239..3490
FT /*tag= a
XX
PN WO9800549-A1.
XX
PD 08-JAN-1998.
XX
PF 24-JUN-1997; 97WO-AU00402.
XX
PR 27-JUN-1996; 96AU-0000699.
XX
PA (CSIR) COMMONWEALTH SCI & IND RES ORG.
PA (AUSU) UNIV AUSTRALIAN NAT.
XX
PI Arioli A, Betzner AS, Peng L, Williamson RE;
XX
DR WPI; 1998-086974/08.
DR P-PSDB; AAW33818.
XX
PT DNA encoding cellulose biosynthetic enzyme - useful for manipulation
PT of cellulose and beta-1,4-glucan
XX
PS Claim 12; Page 123-130; 207pp; English.
XX
CC cDNA clone Ath-A was isolated from an Arabidopsis thaliana cDNA
CC library using probes produced by PCR with primers (see AAT99632-34)
CC based on cellulose synthase RSW1 genomic clone 23H12 (see AAV06563)
CC and EST clone AAT20782 (see AAV06562). It is closely related to RSW1
CC cDNA (see AAV06565) and to partial genomic clone 12C4 (see AAV06564),
CC and encodes a 1084-polypeptide (see AAW33818). Claimed nucleic acid
CC molecules (see AAV06562-69) coding for claimed polypeptides (see
CC AAW33816-20 and AAW46202) involved in cellulose biosynthesis can be
CC used to manipulate the cellulose and/or beta-glucan content of
CC transgenic plants. Expression of nucleic acids in the sense
CC orientation increases the level of cellulose and reduces the level
CC of non-crystalline beta-1,4-glucan and starch, providing plants
CC with modified strength and/or shape and/or fibre properties, or
CC having increased resistance to stresses or pests. Antisense,

C ribozyme or co-suppression molecules can be used to reduce the
C cellulose content of a transgenic plant, e.g. to improve
C digestibility or to alter carbon partitioning such that increased
C carbon is available for growth, rather than deposited as cellulose.
X
Q Sequence 3828 BP; 1056 A; 721 C; 912 G; 1139 T; 0 other;

Query Match 56.2%; Score 1934.4; DB 19; Length 3828;
Best Local Similarity 76.2%; Pred. No. 0;
Matches 2478; Conservative 0; Mismatches 731; Indels 45; Gaps 6;

```
y        8 CGGTTAATCGCCGGTTCTCACAACAGGAATGAGTTTGTCCTCATTAATGCCGATGAGAAT 67
b      254 CGGCTCATTGCTGGCTCTCACAACAGAAACGAATTCGTTCTCATTAACGCCGATGAGAGT 313

y       68 GCCCGAATAAGATCAGTCCAAGAGCTGAGTGGACAGACATGTCAAATCTGCAGAGATGAG 127
b      314 GCCAGAATACGATCAGTACAAGAACTGAGTGGGCAAACATGTCAAATCTGTGGAGATGAA 373

y      128 ATCGAATTGACTGTTGATGGAGAACCGTTTGTGGCATGTAACGAATGTGCATTCCCTGTG 187
b      374 ATCGAATTAACGGTTAGCAGTGAGCTCTTTGTTGCTTGCAACGAATGCGCATTCCCGGTT 433

y      188 TGTAGACCTTGCTATGAGTACGAAAGACGAGAAGGCAATCAAGCTTGTCCACAGTGCAAA 247
b      434 TGTAGACCATGCTATGAGTATGAACGTAGAGAAGGAAATCAAGCTTGTCCTCAGTGCAAA 493

y      248 ACCCGTTTCAAACGTCTTAAAGGAAGTCCAAGAGTTGA---AGGTGATGAAGAGGAAGAT 304
b      494 ACTCGATACAAAAGGATTAAAGGTAGTCCACGGGTTGATGGAGATGATGAAGAAGAAGAA 553

y      305 GACATTGATGATTTAGACAATGAGTTTGAGTATGGAAATAATGGGATTGGATTTGATCAG 364
b      554 GACATTGATGATCTTGAGTATGAGTTTG---------ATCATGGGATGGACCCTGAACAT 604

y      365 GTTTCTGAAGGTATGTCAATCTCTCGTCGCAACTCCGGTTTCCCACAATCTGATTTGGAT 424
b      605 GCCGCTGAAGCCGCACTCTCTTCACGCCTTAACACCGGT------CGTGGTGGATTGGAT 658

y      425 TCAGCTCCACCTGGCTCTCAGATTCCATTGCTGACTTACGGCGACGAGGACGTTGAGATT 484
b      659 TCAGCTCCACCTGGCTCTCAGATTCCTCTTTTGACTTATTGTGATGAAGATGCTGATATG 718

y      485 TCTTCTGATAGACATGCTCTTATTGTTCCTCCTTCACTTGGTGGTCATGGCAATAGAGTT 544
b      719 TATTCTGATCGTCATGCTCTTATCGTGCCTCCTTCA---ACGGGATATGGGAATCGCGTC 775

y      545 CATCCTGTTTCTCTTTCTGACCCGACCGTGGCTGCACATCGAAGGCTGATGGTACCTCAG 604
b      776 TATCCTGCACCGTTTACAGATTCTTCTGCACCTCCACAGGCGAGATCAATGGTTCCTCAG 835

y      605 AAAGATCTTGCGGTTTATGGTTATGGAAGTGTCGCTTGGAAAGATCGGATGGAGGAATGG 664
b      836 AAAGATATTGCGGAATATGGTTATGGAAGTGTTGCTTGGAAGGACCGTATGGAAGTTTGG 895

y      665 AAGAGAAAGCAGAATGAGAAACTTCAGGTTGTTAGGCATGAAGGAG-------------- 710
b      896 AAGAGACGACAAGGCGAAAAGCTTCAAGTCATTAAGCATGAAGGAGGAAACAATGGTCGA 955

y      711 ----ATCCTGATTTTGAAGATGGTGATGATGCTGATTTTCCAATGATGGATGAGGGAAGG 766
b      956 GGTTCCAATGATGACGACGAACTAGATGATCCTGACATGCCTATGATGGATGAAGGAAGA 1015

y      767 CAGCCATTGTCTATGAAGATACCAATCAAATCGAGCAAGATAAATCCTTACCGGATGTTA 826
b     1016 CAACCTCTCTCAAGAAAGCTACCTATTCGTTCAAGCAGAATAAATCCTTACAGGATGTTA 1075

y      827 ATTGTGCTACGTCTTGTGATTCTTGGTCTCTTCTTTCACTACCGTATTCTTCACCCCGTC 886
b     1076 ATTCTGTGTCGCCTCGCGATTCTTGGTCTTTTCTTTCATTATAGAATTCTCCATCCAGTC 1135

y      887 AAAGATGCATATGCTTTGTGGCTTATTTCTGTTATATGTGAGATATGGTTTGCTGTTTCA 946
```

```
Db    1136 AATGATGCATATGGATTATGGTTAACGTCAGTTATATGCGAAATATGGTTTGCAGTGTCT 1195

Qy     947 TGGGTTCTTGATCAGTTCCCTAAATGGTACCCTATCGAGCGAGAAACGTACTTGGACCGA 1006
Db    1196 TGGATTCTTGATCAATTCCCCAAATGGTATCCTATAGAACGTGAAACATACCTCGATAGA 1255

Qy    1007 CTCTCATTAAGATATGAGAAAGAAGGGAAACCGTCGGGACTATCCCCTGTGGATGTATTT 1066
Db    1256 CTCTCTCTCAGGTACGAGAAGGAAGGAAAACCGTCAGGATTAGCACCTGTTGATGTTTTT 1315

Qy    1067 GTTAGTACAGTGGATCCATTGAAAGAGCCTCCGCTTATTACTGCAAATACTGTCTTGTCT 1126
Db    1316 GTTAGTACAGTGGATCCGTTGAAAGAGCCCCCCTTGATTACAGCAAACACAGTTCTTTCC 1375

Qy    1127 ATTCTTGCTGTTGATTATCCTGTCGATAAGGTTGCTTGTTACGTATCTGATGATGGTGCT 1186
Db    1376 ATTCTAGCAGTTGATTATCCTGTGGATAAGGTTGCGTGTTATGTATCAAACAATGGTGCA 1435

Qy    1187 GCTATGCTTACATTTGAAGCTCTCTCTGATACAGCTGATTTTGCTACAAAATGGGTTCCT 1246
Db    1436 GCTATGCTTACATTTGAAGCTCTCTCTGATACAGCTGATTTTGCTACAAAATGGGTTCCT 1495

Qy    1247 TTCTGCAAGAAATATTGTATTGAGCCTCGTGCTCCCGAATGGTATTTCTGCCATAAAATG 1306
Db    1496 TTTTGTAAGAAGTTTAATATCGAGCCACGAGCTCCTGAGTGGTATTTTTCTCAGAAGATG 1555

Qy    1307 GACTACTTGAAGAATAAAGTTCATCCCGCATTTGTTAGGGAGCGGCGAGCCATGAAGAGA 1366
Db    1556 GATTACCTGAAGAACAAAGTTCATCCTGCTTTTGTCAGGGAACGTCGTGCTATGAAGAGA 1615

Qy    1367 GATTATGAAGAATTCAAAGTAAAGATCAATGCTTTAGTAGCAACAGCACAGAAAGTGCCT 1426
Db    1616 GATTATGAAGAGTTTAAAGTGAAGATAAATGCACTGGTTGCTACTGCACAGAAAGTGCCT 1675

Qy    1427 GAGGATGGTTGGACTATGCAAGACGGTACACCTTGGCCCGGTAATAGTGTGCGAGATCAT 1486
Db    1676 GAGGAACGTTGGACTATGCAAGATGGAACTCCTTGGCCTGGAAACAACGTCCGTGACCAT 1735

Qy    1487 CCTGGCATGATTCAGGTCTTCCTTGGAAGTGACGGTGTTCGTGATGTCGAAAACAACGAG 1546
Db    1736 CCTGGAATGATTCAGGTGTTCTTGGGTCATAGTGGAGTTCGTGATACGGATGGTAATGAG 1795

Qy    1547 TTGCCTCGATTAGTTTACGTTTCTCGTGAGAAGAGACCCGGATTTGATCACCATAAGAAG 1606
Db    1796 TTACCACGTCTAGTGTATGTTTCTCGTGAGAAGCGGCCTGGATTTGATCACCACAAGAAA 1855

Qy    1607 GCTGGAGCTATGAATTCCCTGATACGAGTCTCTGGGGTTCTATCAAATGCTCCTTACCTT 1666
Db    1856 GCTGGAGCTATGAATTCCTTGATCCGAGTCTCTGCTGTTCTATCAAACGCTCCTTACCTT 1915

Qy    1667 CTGAATGTCGATTGTGATCACTACATCAACAATAGCAAAGCTCTTAGAGAAGCAATGTGT 1726
Db    1916 CTTAATGTCGATTGTGATCACTACATCAACAACAGCAAAGCAATTAGAGAATCTATGTGT 1975

Qy    1727 TTCATGATGGATCCTCAGTCAGGAAAGAAAATCTGTTATGTTCAGTTCCCTCAAAGGTTC 1786
Db    1976 TTCATGATGGACCCGCAATCGGGAAAGAAAGTTTGTTATGTTCAGTTTCCGCAGAGATTT 2035

Qy    1787 GATGGGATTGATAGGCACGATCGATACTCAAATCGCAATGTTGTGTTCTTTGATATCAAT 1846
Db    2036 GATGGGATTGATAGACATGATAGATACTCAAACCGTAACGTTGTGTTCTTTGATATTAAC 2095

Qy    1847 ATGAAAGGTTTGGATGGGCTACAAGGGCCTATATACGTCGGTACAGGTTGTGTTTTCAGG 1906
Db    2096 ATGAAAGGTCTTGATGGGATACAAGGACCGATATATGTCGGGACAGGTTGTGTGTTTAGA 2155

Qy    1907 AGGCAAGCGCTTTACGGATTTGATGCACCGAAGAAGAAGAAGGGCCCACGTAAGACATGC 1966
Db    2156 AAACAGGCTCTTTATGGTTTTGATGCACCAAAGAAGAAGAAACCACCAGGCAAAACCTGT 2215

Qy    1967 AATTGCTGGCCAAAATGGTGTCTCCTATGTTTTGGTTCAAGAAAGAATCGTAAAGCAAAG 2026
Db    2216 AACTGTTGGCCTAAATGGTGTTGTTTGTGTTGTGGGTTGAGAAAGAAGAGTAAAACGAAA 2275
```

```
Qy       2027 ACAGTGGCTGCGGATAAGAAGAAGAAGAATAGGGAAGCGTCAAAGCAGATCCACGCATTA 2086
                 ||  | |||||||| |  || | ||  ||  | ||||||||||| || ||  ||
Db       2276 ------GCCACAGATAAGAAAACTAACACTAAAGAGACTTCAAAGCAGATTCATGCGCTA 2329

Qy       2087 GAAAATATCGAAGAGGGCCGCGGTCATAAAGTTCTTAACGTAGAACAGTCAACCGAGGCA 2146
              || ||| |||| || ||           |||    || || ||  ||  | | || |||
Db       2330 GAGAATGTCGACGAAGGTGTTATCGTCCCAGTGTCAAATGTTGAGAAGAGATCTGAAGCA 2389

Qy       2147 ATGCAAATGAAGTTGCAGAAGAAATATGGGCAGTCTCCTGTATTTGTTGCATCTGCGCGT 2206
              |  ||| |||| ||| ||||||| | ||| || ||||| || || |||||  |||||   |
Db       2390 ACACAATTGAAATTGGAGAAGAAGTTTGGACAATCTCCGGTTTTCGTTGCCTCTGCTGTT 2449

Qy       2207 CTGGAGAATGGTGGGATGGCTAGAAACGCAAGCCCGGCTTGTCTGCTTAAAGAAGCCATC 2266
              ||  |||| |||||  |  |  | ||||||||||| || || || | | |||||||||
Db       2450 CTACAGAACGGTGGAGTTCCCCGTAACGCAAGCCCCGCATGTTTGTTAAGAGAAGCCATT 2509

Qy       2267 CAAGTCATTAGTCGCGGATATGAAGATAAAACTGAATGGGGAAAAGAGATTGGGTGGATC 2326
              |||||  |||||  ||||  ||  |||||||||| ||||||||||||||||  ||||||||
Db       2510 CAAGTTATTAGCTGCGGGTACCAAGATAAAACCGAATGGGGAAAAGAGATCGGGTGGATT 2569

Qy       2327 TATGGTTCTGTTACCGAAGATATTCTTACGGGTTCTAAGATGCATTCTCATGGTTGGAGA 2386
              |||||  ||  ||  || ||||||||  || |||||||   |||||||||   |||||  ||||||
Db       2570 TATGGATCGGTGACTGAAGATATCCTGACGGGTTTCAAGATGCATTGCCATGGATGGAGA 2629

Qy       2387 CATGTTTATTGTACACCAAAGTTAGCGGCTTTCAAAGGATCAGCTCCAATCAATCTTTCG 2446
                 ||| || ||||  || |||   || |||||  |||||||| |||||  ||  ||  | ||
Db       2630 TCTGTGTACTGTATGCCTAAGCGTGCAGCTTTTAAAGGATCTGCTCCTATTAACTTGTCA 2689

Qy       2447 GATCGTCTCCATCAAGTTCTTCGATGGGCGCTTGGGTCGGTTGAGATTTTCTTGAGTAGG 2506
              ||||||||  ||||||||||  || |||||  |||||  ||  ||  ||||||||||||| ||
Db       2690 GATCGTCTTCATCAAGTTCTACGTTGGGCTCTTGGCTCTGTAGAGATTTTCTTGAGCAGA 2749

Qy       2507 CATTGTCCTATTTGGTATGGTTATGGAGGTGGGTTGAAATGGCTTGAGCGGTTGTCCTAC 2566
              |||||||  ||  |||||||||||||||  |||||  ||  ||||||  |  |||  |  ||  |||
Db       2750 CATTGTCCGATATGGTATGGTTATGGTGGTGGTTTAAAATGGTTGGAGAGATTCTCTTAC 2809

Qy       2567 ATTAACTCTGTGGTTTACCCGTGGACCTCTCTACCGCTCATCGTTTACTGTTCTCTCCCT 2626
              ||  |||||||  ||  ||  ||  |||||  ||  ||  ||  |  |||||  ||  ||||||||||||
Db       2810 ATCAACTCTGTCGTCTATCCTTGGACTTCACTTCCATTGATCGTCTATTGTTCTCTCCCC 2869

Qy       2627 GCCATCTGTCTTCTCACTGGAAAATTCATCGTTCCCGAGATTAGCAACTATGCGAGTATC 2686
              ||  |  |||  |  |||||  ||||||||||||||  ||  |||||  ||||||||  ||  ||||
Db       2870 GCGGTTTGTTTACTCACAGGAAAATTCATCGTCCCTGAGATAAGCAACTACGCAGGTATA 2929

Qy       2687 CTCTTCATGGCGCTCTTCTCGTCGATTGCAATAACGGGTATTGTCGAGATGCAATGGGGC 2746
              |||||||||     | |||   ||  ||  ||  |||| ||  ||  ||| ||||||||||
Db       2930 CTCTTCATGCTCATGTTCATATCGATAGCAGTAACTGGAATTGTCGAAATGCAATGGGGA 2989

Qy       2747 AAAGTTGGGATCGATGATTGGTGGAGAAACGAACAGTTTTGTGTCATTGGAGGTGTTTCT 2806
                 ||  ||  ||||||||||||||||||||  |||||  |||||||||  ||  |||||  |  ||
Db       2990 GGTGTCGGAATCGATGATTGGTGGAGAAACGAGCAGTTTTGGGTAATCGGAGGGGCCTCC 3049

Qy       2807 GCGCATCTGTTTGCTCTCTTCCAAGGTCTCCTCAAGGTTCTTGCTGGTGTCGACACTAAC 2866
              |||||||  ||||||||  ||  ||||||  |  |||||  |||||  ||  ||  ||||  ||
Db       3050 TCGCATCTATTTGCTCTGTTTCAAGGTTTGCTCAAAGTTCTAGCCGGAGTTAACACGAAT 3109

Qy       2867 TTCACAGTCACATCAAAAGCAGCTGATGATGGAGAGTTCTCTGACCTTTACCTCTTCAAA 2926
              ||||||||||  ||||||||||  ||  ||||||  |||||||  ||||||  |||||||
Db       3110 TTCACAGTCACTTCAAAAGCAGCAGACGATGGAGCTTTCTCTGAGCTTTACATCTTCAAG 3169

Qy       2927 TGGACTTCACTTCTCATCCCTCCAATGACTCTACTCATCATAAACGTCATTGGAGTCATA 2986
              |||||  |  |  | ||  |||||  |  ||  ||  ||||  |||  ||||||||||  ||
Db       3170 TGGACAACTTTGTTGATTCCTCCGACAACACTTCTGATCATTAACATCATTGGAGTTATT 3229

Qy       2987 GTCGGAGTCTCTGATGCCATCAGCAATGGATACGACTCGTGGGGACCGCTTTTCGGAAGA 3046
              |||||  ||  ||||||||||  ||||||||  ||  |||||  ||  |||||  ||  ||  |||
Db       3230 GTCGGCGTTTCTGATGCCATTAGCAATGGCTATGACTCATGGGGACCTCTCTTTGGGAGA 3289

Qy       3047 CTGTTCTTTGCACTTTGGGTCATCATTCATCTTTACCCGTTGCTTAAAGGTTTGCTTGGG 3106
              ||  |||||  ||  ||||||||||  |||||  |  |||||  |||||  ||  ||  ||||||||
Db       3290 CTTTTCTTCGCTCTTTGGGTCATTGTTCATTTATACCCATTCCTCAAGGGAATGCTTGGG 3349

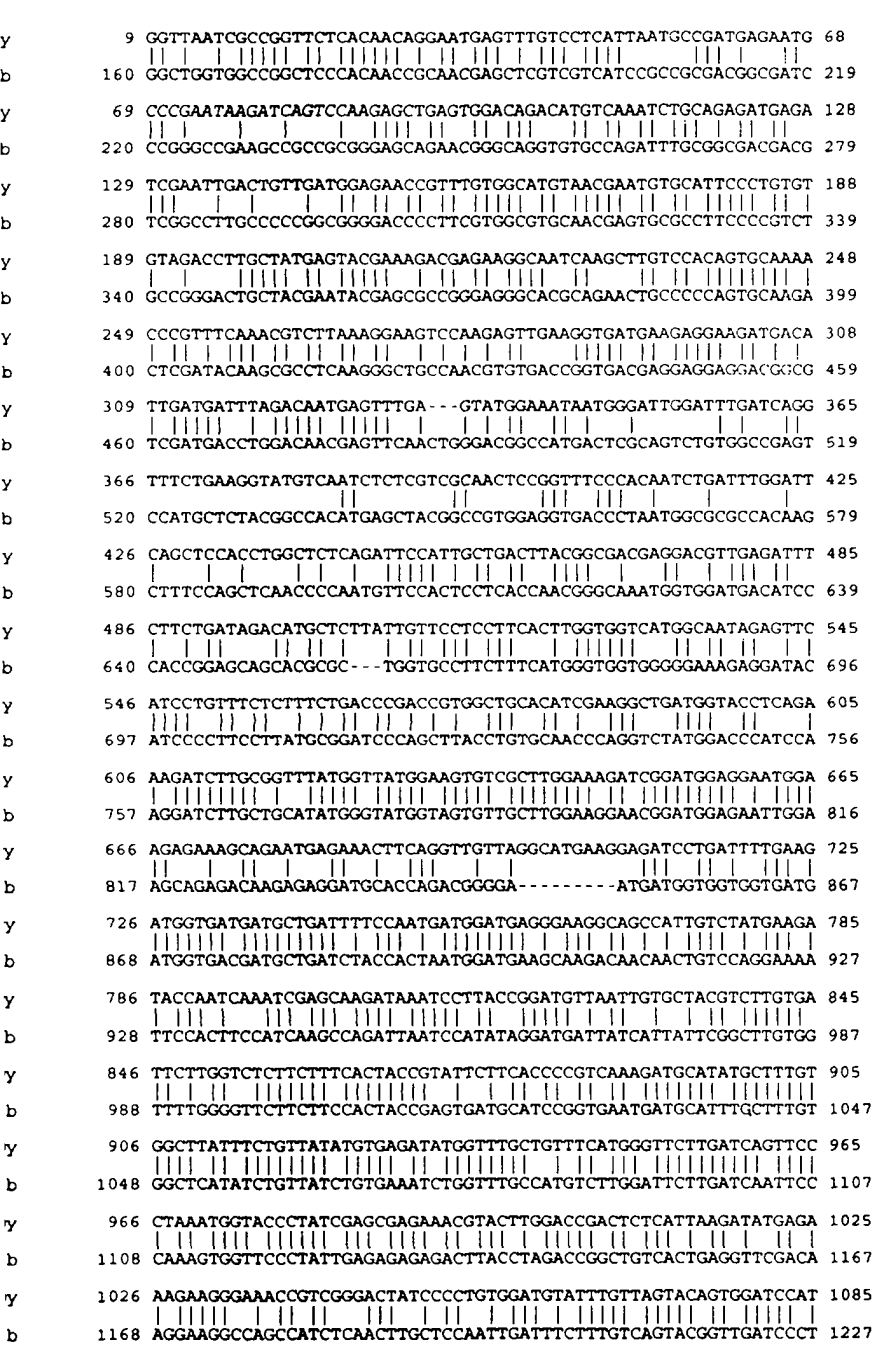
Qy       3107 AAACAAGATAGAATGCCAACCATTATTGTCGTCTGGTCCATCCTCCTGGCCTCGATTCTT 3166
              ||  |||||  |  ||||||  ||  ||||||||  ||||||||  ||  ||  ||  |||||  |
Db       3350 AAGCAAGACAAAATGCCTACGATTATTGTGGTCTGGTCTATTCTTCTAGCTTCGATCTTG 3409

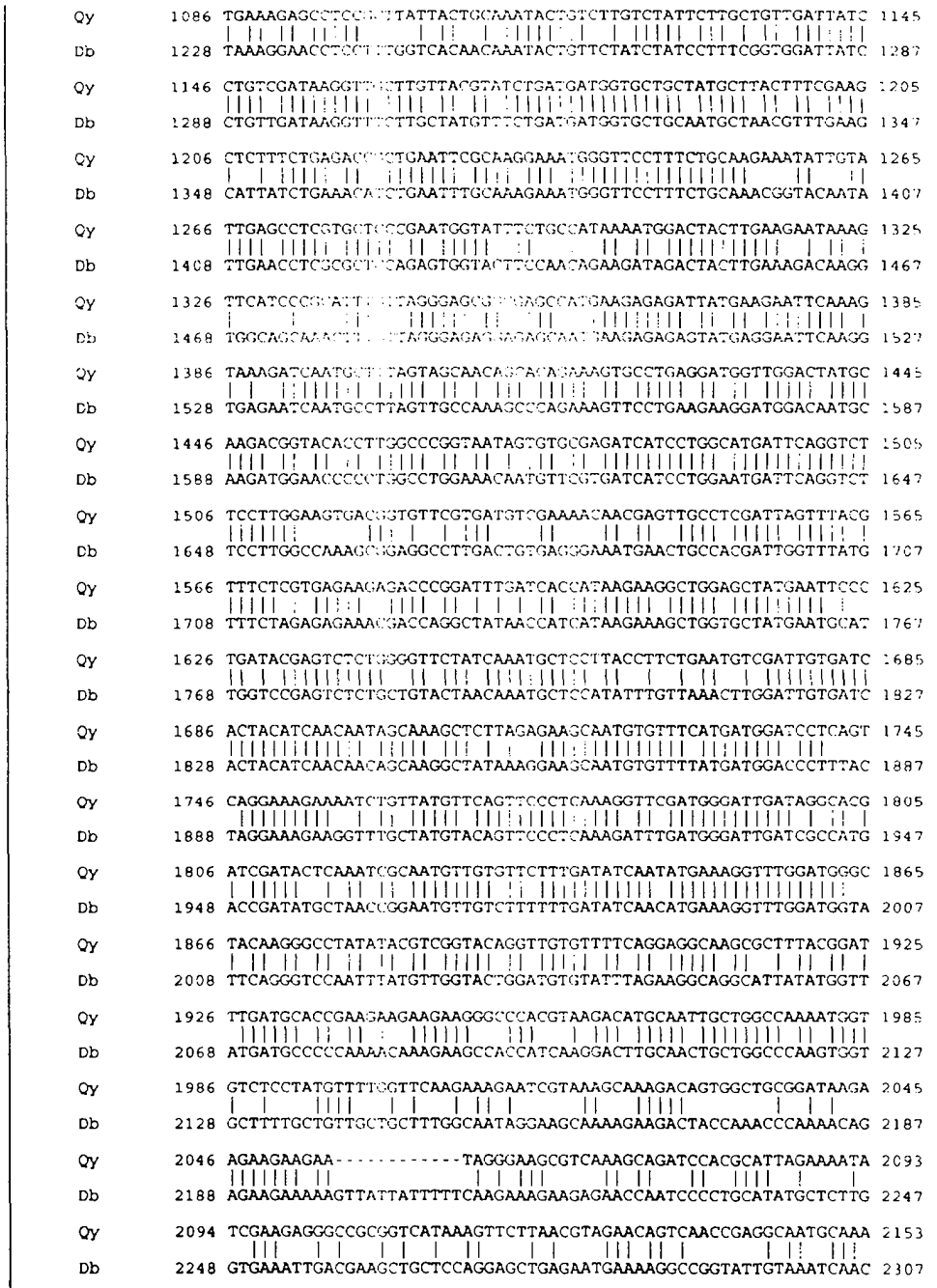
Qy       3167 ACACTTCTTTGGGTCCGGGTTAATCCGTTTGTGGCGAAAGGCGGTCCTATTCTCGAGATC 3226
              |||||  |  ||||||  |  ||||  ||||||||||||  |||||  ||  ||  |  |  ||||||
Db       3410 ACACTCTTGTGGGTCAGAATTAACCCGTTTGTGGCTAAAGGGGGACCAGTGTTGGAGATC 3469

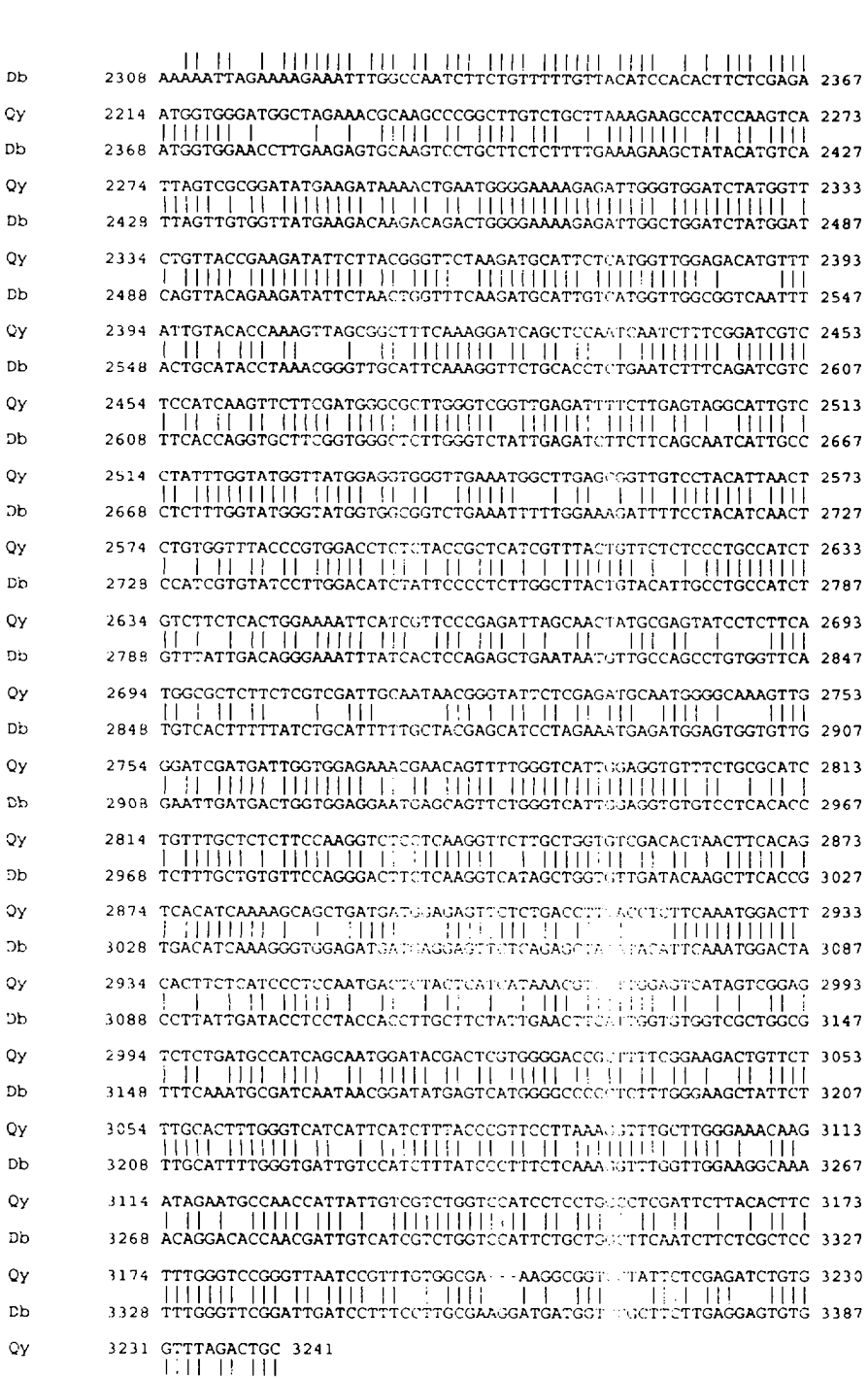
Qy       3227 TGTGGTTTAGACTG 3240
              ||||||  |  | ||
Db       3470 TGTGGTCTGAATTG 3483
```

RESULT 5
AAZ58266
ID AAZ58266 standard; cDNA; 3936 BP.
XX
AC AAZ58266;
XX
DT 08-MAY-2000 (first entry)
XX
DE Corn cellulose synthase cDNA clone p0122.ckamh70rc.
XX
KW Corn; maize; cellulose synthase; transgenic plant; ss.
XX
OS Zea mays.
XX
FH Key Location/Qualifiers
FT CDS 144..3404
FT /*tag= a
XX
PN WO200004166-A2.
XX
PD 27-JAN-2000.
XX
PF 13-JUL-1999; 99WO-US15871.
XX
PR 14-JUL-1998; 98US-0092844.
XX
PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX
PI Allen SM, Fader GM, Falco SC, Kinney AJ, Lightner JE, Miao G;
PI Rafalski JA, Thorpe CJ;
XX
DR WPI; 2000-182431/16.
DR P-PSDB; AAY58835.
XX
PT Nucleic acid fragments encoding cellulose biosynthetic enzyme useful as
PT probes for isolating cDNAs and genes encoding homologous proteins, for
PT producing transgenic plants -
XX
PS Claim 9; Page 49-50; 93pp; English.
XX
CC The present sequence is that of full-length cDNA clone
CC p0122.ckamh70rc encoding corn cellulose synthase (CS, see AAY58835).
CC The cDNA clone was isolated from a cDNA library produced from corn
CC pith tissue collected from internode subtending ear node 5 days
CC after pollination. The invention relates to isolated nucleic acid
CC fragments encoding plant CS and to CS polypeptides. It also relates
CC to the construction of a chimeric gene encoding all or a portion of
CC the CS, in sense or antisense orientation, where expression of the
CC gene results in altered levels of the CS in transformed host cells.
CC The host cells can be used to screen compounds for their ability to
CC inhibit CS activity. CS nucleic acids are also useful for producing
CC transgenic plants having altered levels of CS, and hence altered
CC levels of fibre. CS may also serve as a target for the development
CC of novel herbicides.
XX
SQ Sequence 3936 BP; 989 A; 838 C; 1018 G; 1091 T; 0 other;

Query Match 43.3%; Score 1489.8; DB 21; Length 3936;
Best Local Similarity 67.3%; Pred. No. 0;
Matches 2189; Conservative 0; Mismatches 1032; Indels 30; Gaps 5;



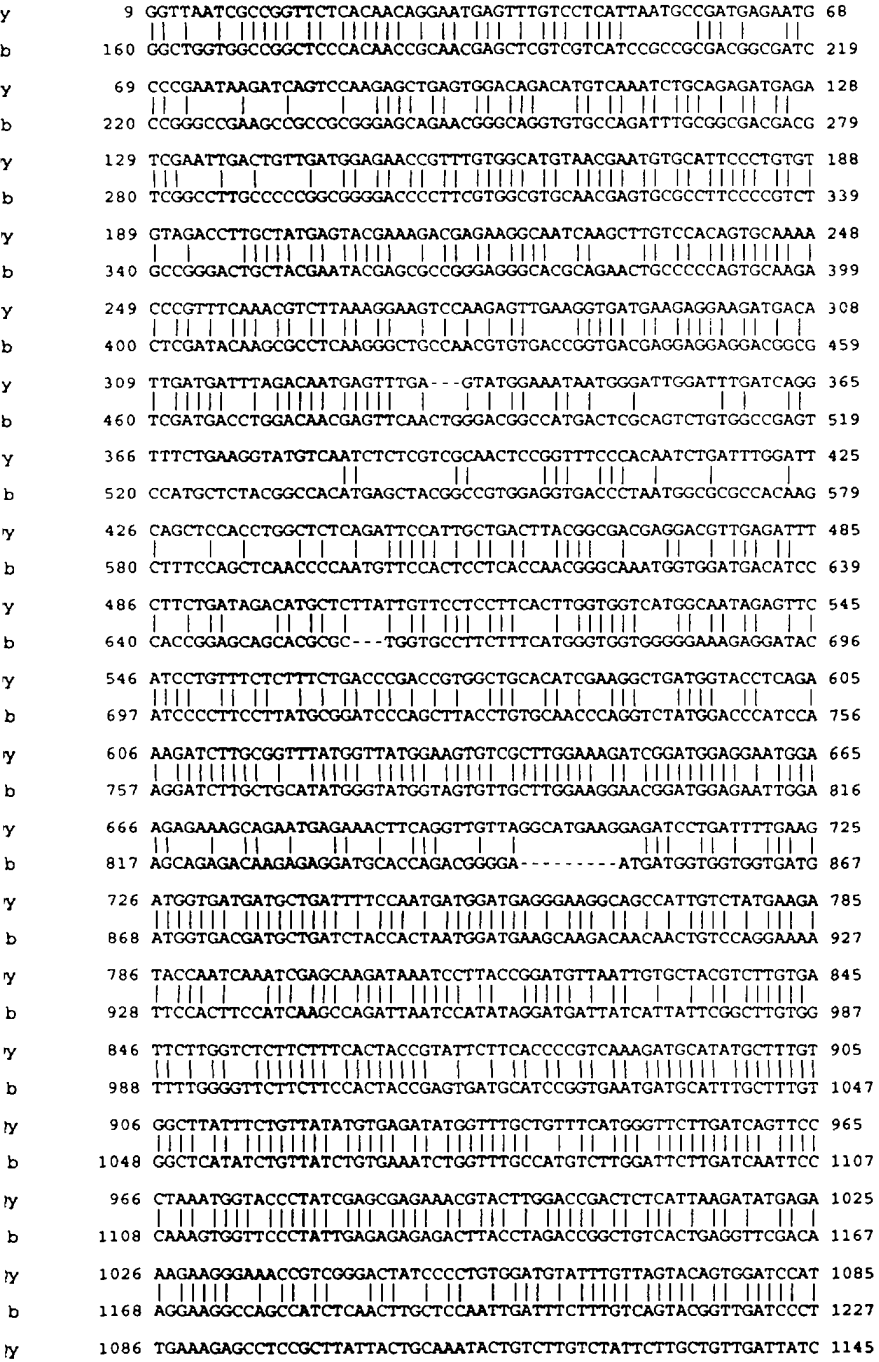




Db 3388 GTTTGGATTGC 3398

RESULT 6
AAZ99503
ID AAZ99503 standard; DNA; 3969 BP.
XX
AC AAZ99503;
XX
DT 03-JUL-2000 (first entry)
XX
DE DNA encoding a maize cellulose synthase.
XX
KW Maize; cellulose synthase; stalk quality; stand; silage; cellulose;
KW transgenic plant; plant breeding marker; ss.
XX
OS Zea mays.
XX
FH Key Location/Qualifiers
FT CDS 144..3398
FT /*tag= a
FT /product= "cellulose synthase"
FT /note= "no termination codon given"
XX
PN WO200009706-A2.
XX
PD 24-FEB-2000.
XX
PF 16-AUG-1999; 99WO-US18760.
XX
PR 17-AUG-1998; 98US-0096822.
XX
PA (PION-) PIONEER HI-BRED INT INC.
XX
PI Dhugga KS, Helentjaris TG, Bowen BA, Wang X;
XX
DR WPI; 2000-224343/19.
DR P-PSDB; AAY84111.
XX
PT New genes which encode maize cellulose synthase polypeptides in plants
PT useful for modulating the expression of cellulose synthase in plants
PT and to produce transgenic plants expressing the novel protein -
XX
PS Claim 1; Page 113-118; 119pp; English.
XX
CC The present sequence encodes a maize cellulose synthase polypeptide.
CC The cellulose synthase can be used for the improvement of stalk quality
CC for improved stand or silage. It also provides an increased concentration
CC of cellulose in the pericarp, hardening the kernel and improving its
CC handling ability. The sequences are used to produce transgenic plants
CC and seeds expressing the cellulose synthase. The polynucleotide is
CC used for modulating, preferably increasing, the level of the synthase
CC in a plant cell. The plants are preferably monocots. The polynucleotide
CC is also used as a probe or primer in the detection quantitation or
CC isolation of gene transcripts. The probes are useful in detecting
CC deficiencies in the level of mRNA in screenings for desired transgenic
CC plant, for detecting mutations in the gene, for monitoring upregulation
CC of expression or changes in enzyme activity in screening assays of
CC compounds, for detection of any number of allelic variants of the gene,
CC or for use as molecular markers in plant breeding programs. The
CC isolated nucleic acids of the present invention can also be used for
CC recombinant expression of their encoded polypeptides or for use as
CC immunogens in the preparation and/or screening of antibodies. The
CC proteins can be employed in assays for enzyme agonists or antagonists
CC of enzyme function or for use of immunogens or antigens to obtain
CC antibodies specifically immunoreactive with a protein.
XX
SQ Sequence 3969 BP; 1022 A; 838 C; 1018 G; 1091 T; 0 other;

Query Match 43.3%; Score 1489.8; DB 21; Length 3969;
Best Local Similarity 67.3%; Pred. No. 0;
Matches 2189; Conservative 0; Mismatches 1032; Indels 30; Gaps 5;


y 9 GGTTAATCGCCGGTTCTCACAACAGGAATGAGTTTGTCCTCATTAATGCCGATGAGAATG 68
b 160 GGCTGGTGGCCGGCTCCCACAACCGCAACGAGCTCGTCGTCATCCGCCGCGACGGCGATC 219
y 69 CCCGAATAAGATCAGTCCAAGAGCTGAGTGGACAGACATGTCAAATCTGCAGAGATGAGA 128
b 220 CCGGGCCGAAGCCGCCGCGGGAGCAGAACGGGCAGGTGTGCCAGATTTGCGGCGACGACG 279
y 129 TCGAATTGACTGTTGATGGAGAACCGTTTGTGGCATGTAACGAATGTGCATTCCCTGTGT 188
b 280 TCGGCCTTGCCCCCGGCGGGGACCCCTTCGTGGCGTGCAACGAGTGCGCCTTCCCCGTCT 339
y 189 GTAGACCTTGCTATGAGTACGAAAGACGAGAAGGCAATCAAGCTTGTCCACAGTGCAAAA 248
b 340 GCCGGGACTGCTACGAATACGAGCGCCGGGAGGGCACGCAGAACTGCCCCCAGTGCAAGA 399
y 249 CCCGTTTCAAACGTCTTAAAGGAAGTCCAAGAGTTGAAGGTGATGAAGAGGAAGATGACA 308
b 400 CTCGATACAAGCGCCTCAAGGGCTGCCAACGTGTGACCGGTGACGAGGAGGAGGACGGCG 459
y 309 TTGATGATTTAGACAATGAGTTTGA---GTATGGAAATAATGGGATTGGATTTGATCAGG 365
b 460 TCGATGACCTGGACAACGAGTTCAACTGGGACGGCCATGACTCGCAGTCTGTGGCCGAGT 519
y 366 TTTCTGAAGGTATGTCAATCTCTCGTCGCAACTCCGGTTTCCCACAATCTGATTTGGATT 425
b 520 CCATGCTCTACGGCCACATGAGCTACGGCCGTGGAGGTGACCCTAATGGCGCGCCACAAG 579
y 426 CAGCTCCACCTGGCTCTCAGATTCCATTGCTGACTTACGGCGACGAGGACGTTGAGATTT 485
b 580 CTTTCCAGCTCAACCCCAATGTTCCACTCCTCACCAACGGGCAAATGGTGGATGACATCC 639
y 486 CTTCTGATAGACATGCTCTTATTGTTCCTCCTTCACTTGGTGGTCATGGCAATAGAGTTC 545
b 640 CACCGGAGCAGCACGCGC---TGGTGCCTTCTTTCATGGGTGGTGGGGGAAAGAGGATAC 696
y 546 ATCCTGTTTCTCTTTCTGACCCGACCGTGGCTGCACATCGAAGGCTGATGGTACCTCAGA 605
b 697 ATCCCCTTCCTTATGCGGATCCCAGCTTACCTGTGCAACCCAGGTCTATGGACCCATCCA 756
y 606 AAGATCTTGCGGTTTATGGTTATGGAAGTGTCGCTTGGAAAGATCGGATGGAGGAATGGA 665
b 757 AGGATCTTGCTGCATATGGGTATGGTAGTGTTGCTTGGAAGGAACGGATGGAGAATTGGA 816
y 666 AGAGAAAGCAGAATGAGAAACTTCAGGTTGTTAGGCATGAAGGAGATCCTGATTTTGAAG 725
b 817 AGCAGAGACAAGAGAGGATGCACCAGACGGGGA---------ATGATGGTGGTGGTGATG 867
y 726 ATGGTGATGATGCTGATTTTCCAATGATGGATGAGGGAAGGCAGCCATTGTCTATGAAGA 785
b 868 ATGGTGACGATGCTGATCTACCACTAATGGATGAAGCAAGACAACAACTGTCCAGGAAAA 927
y 786 TACCAATCAAATCGAGCAAGATAAATCCTTACCGGATGTTAATTGTGCTACGTCTTGTGA 845
b 928 TTCCACTTCCATCAAGCCAGATTAATCCATATAGGATGATTATCATTATTCGGCTTGTGG 987
y 846 TTCTTGGTCTCTTCTTTCACTACCGTATTCTTCACCCCGTCAAAGATGCATATGCTTTGT 905
b 988 TTTTGGGGTTCTTCTTCCACTACCGAGTGATGCATCCGGTGAATGATGCATTTGCTTTGT 1047
y 906 GGCTTATTTCTGTTATATGTGAGATATGGTTTGCTGTTTCATGGGTTCTTGATCAGTTCC 965
b 1048 GGCTCATATCTGTTATCTGTGAAATCTGGTTTGCCATGTCTTGGATTCTTGATCAATTCC 1107
y 966 CTAAATGGTACCCTATCGAGCGAGAAACGTACTTGGACCGACTCTCATTAAGATATGAGA 1025
b 1108 CAAAGTGGTTCCCTATTGAGAGAGAGACTTACCTAGACCGGCTGTCACTGAGGTTCGACA 1167
y 1026 AAGAAGGGAAACCGTCGGGACTATCCCCTGTGGATGTATTTGTTAGTACAGTGGATCCAT 1085
b 1168 AGGAAGGCCAGCCATCTCAACTTGCTCCAATTGATTTCTTTGTCAGTACGGTTGATCCCT 1227
y 1086 TGAAAGAGCCTCCGCTTATTACTGCAAATACTGTCTTGTCTATTCTTGCTGTTGATTATC 1145
Db 1228 TAAAGGAACCTCCTTTGGTCACAACAAATACTGTTCTATCTATCCTTTCGGTGGATTATC 1287
Qy 1146 CTGTCGATAAGGTTGCTTGTTACGTATCTGATGATGGTGCTGCTATGCTTACTTTCGAAG 1205
Db 1288 CTGTTGATAAGGTTTCTTGCTATGTTTCTGATGATGGTGCTGCAATGCTAACGTTTGAAG 1347
Qy 1206 CTCTTTCTGAGACCGCTGAATTCGCAAGGAAATGGGTTCCTTTCTGCAAGAAATATTGTA 1265
Db 1348 CATTATCTGAAACATCTGAATTTGCAAAGAAATGGGTTCCTTTCTGCAAACGGTACAATA 1407
Qy 1266 TTGAGCCTCGTGCTCCCGAATGGTATTTCTGCCATAAAATGGACTACTTGAAGAATAAAG 1325
Db 1408 TTGAACCTCGCGCTCCAGAGTGGTACTTCCAACAGAAGATAGACTACTTGAAAGACAAGG 1467
Qy 1326 TTCATCCCGCATTTTTAGGGAGCGGCGAGCCATGAAGAGAGATTATGAAGAATTCAAAG 1385
Db 1468 TGGCAGCAAACTTTTTAGGGAGAGAGAGCAATGAAGAGAGAGTATGAGGAATTCAAGG 1527
Qy 1386 TAAAGATCAATGGTTTAGTAGCAACAGCACAGAAAGTGCCTGAGGATGGTTGGACTATGC 1445
Db 1528 TGAGAATCAATGCCTTAGTTGCCAAAGCCCAGAAAGTTCCTGAAGAAGGATGGACAATGC 1587
Qy 1446 AAGACGGTACACCTTGGCCCGGTAATAGTGTGCGAGATCATCCTGGCATGATTCAGGTCT 1505
Db 1588 AAGATGGAACCCCCTGGCCTGGAAACAATGTTCGTGATCATCCTGGAATGATTCAGGTCT 1647
Qy 1506 TCCTTGGAAGTGACGGTGTTCGTGATGTCGAAAACAACGAGTTGCCTCGATTAGTTTACG 1565
Db 1648 TCCTTGGCCAAAGCGGAGGCCTTGACTGTGAGGGAAATGAACTGCCACGATTGGTTTATG 1707
Qy 1566 TTTCTCGTGAGAAGAGACCCGGATTTGATCACCATAAGAAGGCTGGAGCTATGAATTCCC 1625
Db 1708 TTTCTAGAGAGAAACGACCAGGCTATAACCATCATAAGAAAGCTGGTGCTATGAATGCAT 1767
Qy 1626 TGATACGAGTCTCTGGGGTTCTATCAAATGCTCCTTACCTTCTGAATGTCGATTGTGATC 1685
Db 1768 TGGTCCGAGTCTCTGCTGTACTAACAAATGCTCCATATTTGTTAAACTTGGATTGTGATC 1827
Qy 1686 ACTACATCAACAATAGCAAAGCTCTTAGAGAAGCAATGTGTTTCATGATGGATCCTCAGT 1745
Db 1828 ACTACATCAACAACAGCAAGGCTATAAAGGAAGCAATGTGTTTTATGATGGACCCTTTAC 1887
Qy 1746 CAGGAAAGAAAATCTGTTATGTTCAGTTCCCTCAAAGGTTCGATGGGATTGATAGGCACG 1805
Db 1888 TAGGAAAGAAGGTTTGCTATGTACAGTTCCCTCAAAGATTTGATGGGATTGATCGCCATG 1947
Qy 1806 ATCGATACTCAAATCGCAATGTTGTGTTCTTTGATATCAATATGAAAGGTTTGGATGGGC 1865
Db 1948 ACCGATATGCTAACCGGAATGTTGTCTTTTTTGATATCAACATGAAAGGTTTGGATGGTA 2007
Qy 1866 TACAAGGGCCTATATACGTCGGTACAGGTTGTGTTTTCAGGAGGCAAGCGCTTTACGGAT 1925
Db 2008 TTCAGGGTCCAATTTATGTTGGTACTGGATGTGTATTTAGAAGGCAGGCATTATATGGTT 2067
Qy 1926 TTGATGCACCGAAGAAGAAGAAGGGCCCACGTAAGACATGCAATTGCTGGCCAAAATGGT 1985
Db 2068 ATGATGCCCCCAAAACAAAGAAGCCACCATCAAGGACTTGCAACTGCTGGCCCAAGTGGT 2127
Qy 1986 GTCTCCTATGTTTTGGTTCAAGAAAGAATCGTAAAGCAAAGACAGTGGCTGCGGATAAGA 2045
Db 2128 GCTTTTGCTGTTGCTGCTTTGGCAATAGGAAGCAAAAGAAGACTACCAAACCCAAAACAG 2187
Qy 2046 AGAAGAAGAA----------TAGGGAAGCGTCAAAGCAGATCCACGCATTAGAAAATA 2093
Db 2188 AGAAGAAAAAAGTTATTATTTTTCAAGAAAGAAGAGAACCAATCCCCTGCATATGCTCTTG 2247
Qy 2094 TCGAAGAGGGCCGCGGTCATAAAGTTCTTAACGTAGAACAGTCAACCGAGGCAATGCAAA 2153
Db 2248 GTGAAATTGACGAAGCTGCTCCAGGAGCTGAGAATGAAAAGGCCGGTATTGTAAATCAAC 2307
Qy 2154 TGAAGTTGCAGAAGAAATATGGGCAGTCTCCTGTATTTGTTGCATCTGCGCGTCTGGAGA 2213

```
Db       2308 AAAAATTAGAAAAGAAATTTGGCCAATCTTCTGTTTTTGTTACATCCACACTTCTCGAGA 2367

Qy       2214 ATGGTGGGATGGCTAGAAACGCAAGCCCGGCTTGTCTGCTTAAAGAAGCCATCCAAGTCA 2273

Db       2368 ATGGTGGAACCTTGAAGAGTGCAAGTCCTGCTTCTCTTTTGAAAGAAGCTATACATGTCA 2427

Qy       2274 TTAGTCGCGGATATGAAGATAAAACTGAATGGGGAAAAGAGATTGGGTGGATCTATGGTT 2333

Db       2428 TTAGTTGTGGTTATGAAGACAAGACAGACTGGGGAAAAGAGATTGGCTGGATCTATGGAT 2487

Qy       2334 CTGTTACCGAAGATATTCTTACGGGTTCTAAGATGCATTCTCATGGTTGGAGACATGTTT 2393

Db       2488 CAGTTACAGAAGATATTCTAACTGGTTTCAAGATGCATTGTCATGGTTGGCGGTCAATTT 2547

Qy       2394 ATTGTACACCAAAGTTAGCGGCTTTCAAAGGATCAGCTCCAATCAATCTTTCGGATCGTC 2453

Db       2548 ACTGCATACCTAAACGGGTTGCATTCAAAGGTTCTGCACCTCTGAATCTTTCAGATCGTC 2607

Qy       2454 TCCATCAAGTTCTTCGATGGGCGCTTGGGTCGGTTGAGATTTTCTTGAGTAGGCATTGTC 2513

Db       2608 TTCACCAGGTGCTTCGGTGGGCTCTTGGGTCTATTGAGATCTTCTTCAGCAATCATTGCC 2667

Qy       2514 CTATTTGGTATGGTTATGGAGGTGGGTTGAAATGGCTTGAGCGGTTGTCCTACATTAACT 2573

Db       2668 CTCTTTGGTATGGGTATGGTGGCGGTCTGAAATTTTTGGAAAGATTTTCCTACATCAACT 2727

Qy       2574 CTGTGGTTTACCCGTGGACCTCTCTACCGCTCATCGTTTACTGTTCTCTCCCTGCCATCT 2633

Db       2728 CCATCGTGTATCCTTGGACATCTATTCCCCTCTTGGCTTACTGTACATTGCCTGCCATCT 2787

Qy       2634 GTCTTCTCACTGGAAAATTCATCGTTCCCGAGATTAGCAACTATGCGAGTATCCTCTTCA 2693

Db       2788 GTTTATTGACAGGGAAATTTATCACTCCAGAGCTGAATAATGTTGCCAGCCTGTGGTTCA 2847

Qy       2694 TGGCGCTCTTCTCGTCGATTGCAATAACGGGTATTCTCGAGATGCAATGGGGCAAAGTTG 2753

Db       2848 TGTCACTTTTTATCTGCATTTTTGCTACGAGCATCCTAGAAATGAGATGGAGTGGTGTTG 2907

Qy       2754 GGATCGATGATTGGTGGAGAAACGAACAGTTTTGGGTCATTGGAGGTGTTTCTGCGCATC 2813

Db       2908 GAATTGATGACTGGTGGAGGAATGAGCAGTTCTGGGTCATTGGAGGTGTGTCCTCACACC 2967

Qy       2814 TGTTTGCTCTCTTCCAAGGTCTCCTCAAGGTTCTTGCTGGTGTCGACACTAACTTCACAG 2873

Db       2968 TCTTTGCTGTGTTCCAGGGACTTCTCAAGGTCATAGCTGGTGTTGATACAAGCTTCACCG 3027

Qy       2874 TCACATCAAAAGCAGCTGATGATGGAGAGTTCTCTGACCTTTACCTCTTCAAATGGACTT 2933

Db       3028 TGACATCAAAGGGTGGAGATGATCAGGAGTTCTCAGAGCTATATACATTCAAATGGACTA 3087

Qy       2934 CACTTCTCATCCCTCCAATGACTCTATTCACCATAAACGTCATTGGAGTCATAGTCGGAG 2993

Db       3088 CCTTATTGATACCTCCTACCACCTTGCTTCTATTGAACTTCATTGGTGTGGTCGCTGGCG 3147

Qy       2994 TCTCTGATGCCATCAGCAATGGATACGACTCGTGGGGACCGCTTTTCGGAAGACTGTTCT 3053

Db       3148 TTTCAAATGCGATCAATAACGGATATGAGTCATGGGGCCCCCTCTTTGGGAAGCTATTCT 3207

Qy       3054 TTGCACTTTGGGTCATCATTCATCTTTACCCGTTCCTTAAAGGTTTGCTTGGGAAACAAG 3113

Db       3208 TTGCATTTTGGGTGATTGTCCATCTTTATCCCTTTCTCAAAGGTTTGGTTGGAAGGCAAA 3267

Qy       3114 ATAGAATGCCAACCATTATTGTCGTCTGGTCCATCCTCCTGGCCTCGATTCTTACACTTC 3173

Db       3268 ACAGGACACCAACGATTGTCATCGTCTGGTCCATTCTGCTGGCTTCAATCTTCTCGCTCC 3327

Qy       3174 TTTGGGTCCGGGTTAATCCGTTTGTGGCGA---AAGGCGGTCCTATTCTCGAGATCTGTG 3230

Db       3328 TTTGGGTTCGGATTGATCCTTTCCTTGCGAAGGATGATGGTCCGCTTCTTGAGGAGTGTG 3387

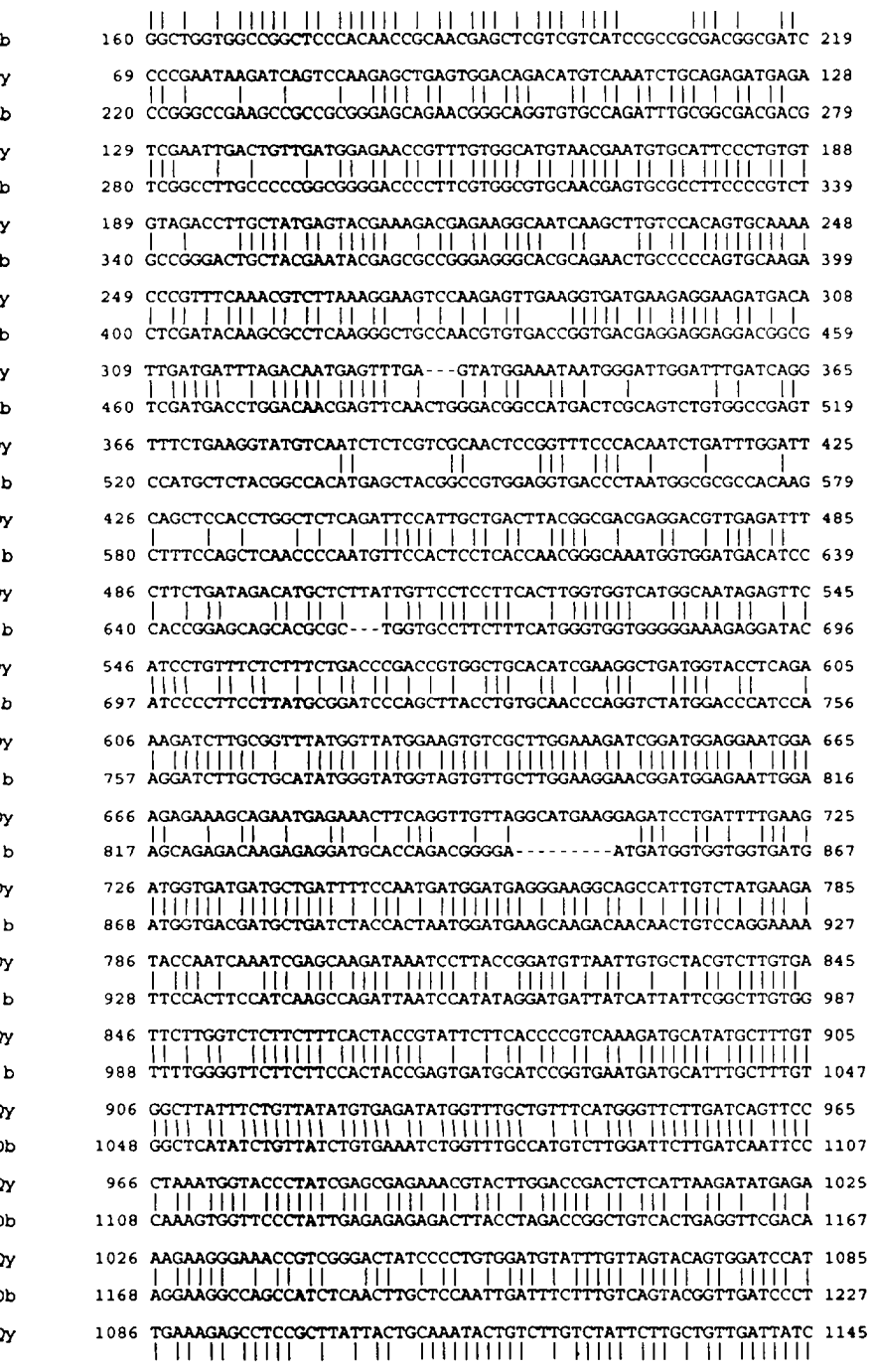
Qy       3231 GTTTAGACTGC 3241

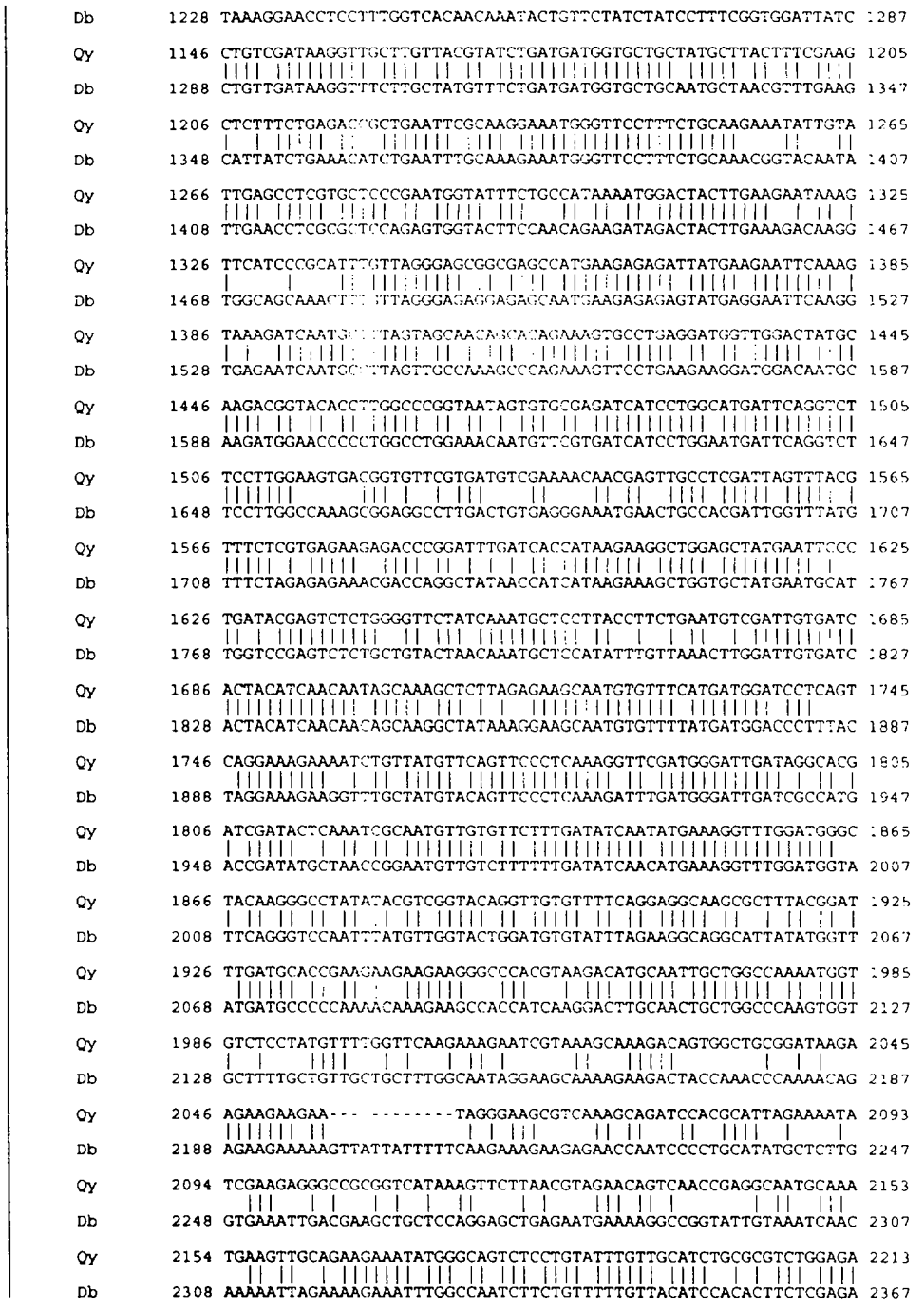
Db       3388 GTTTGGATTGC 3398
```

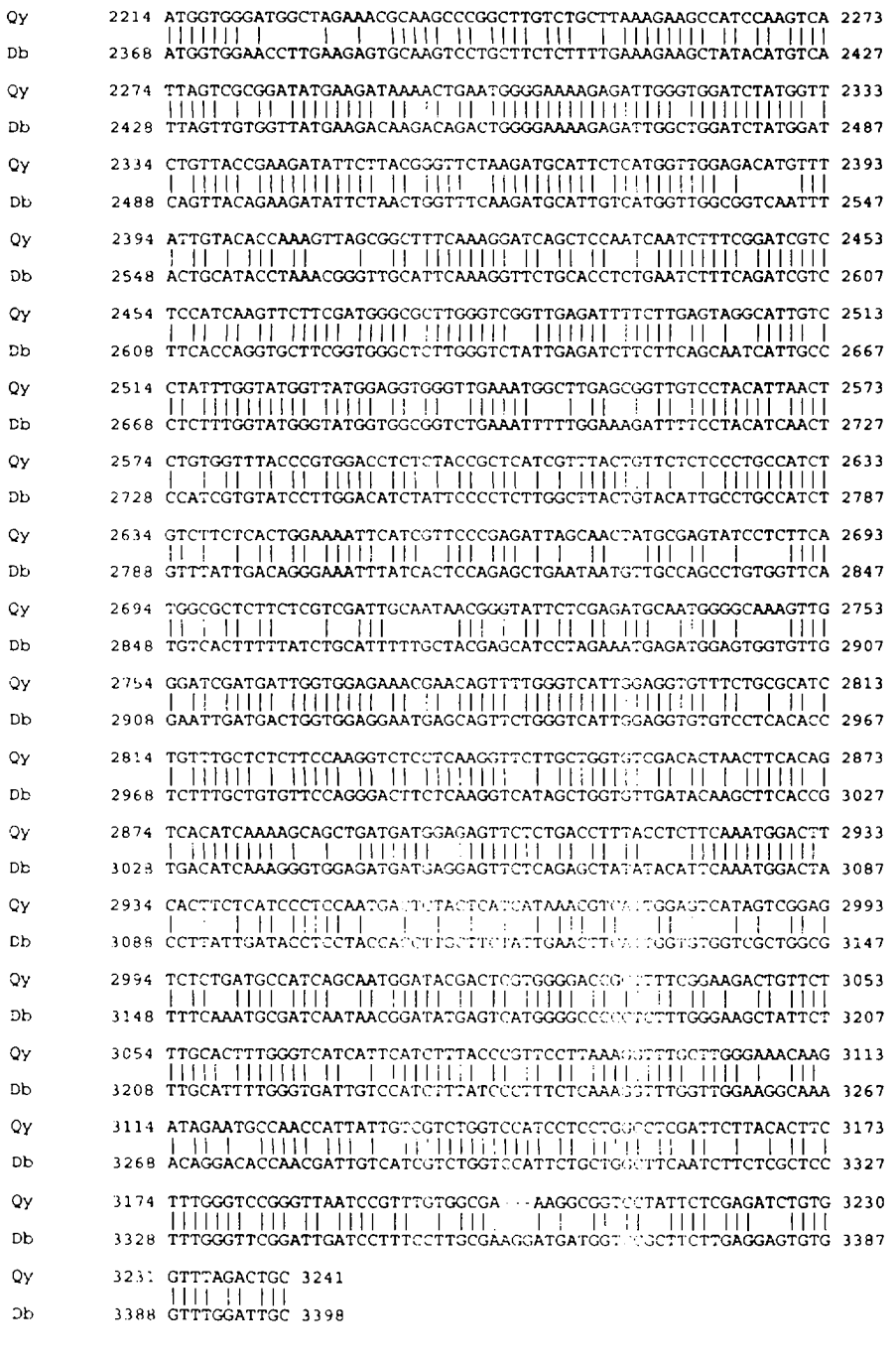
RESULT 7
AAZ99518
ID AAZ99518 standard; DNA; 3969 BP.
XX
AC AAZ99518;
XX
DT 03-JUL-2000 (first entry)
XX
DE DNA encoding a maize cellulose synthase.
XX
KW Maize; cellulose synthase; stalk quality; stand; silage; cellulose;
KW transgenic plant; plant breeding marker; ss.
XX
OS Zea mays.
XX
FH Key Location/Qualifiers
FT CDS 144..3398
FT /*tag= a
FT /product= "cellulose synthase"
FT /note= "no termination codon given"
XX
PN WO200009706-A2.
XX
PD 24-FEB-2000.
XX
PF 16-AUG-1999; 99WO-US18760.
XX
PR 17-AUG-1998; 98US-0096822.
XX
PA (PION-) PIONEER HI-BRED INT INC.
XX
PI Dhugga KS, Helentjaris TG, Bowen BA, Wang X;
XX
DR WPI; 2000-224343/19.
DR P-PSDB; AAY84116.
XX
PT New genes which encode maize cellulose synthase polypeptides in plants
PT useful for modulating the expression of cellulose synthase in plants
PT and to produce transgenic plants expressing the novel protein -
XX
PS Claim 1; Page 152-157; 119pp; English.
XX
CC The present sequence encodes a maize cellulose synthase polypeptide.
CC The cellulose synthase can be used for the improvement of stalk quality
CC for improved stand or silage. It also provides an increased concentration
CC of cellulose in the pericarp, hardening the kernel and improving its
CC handling ability. The sequences are used to produce transgenic plants
CC and seeds expressing the cellulose synthase. The polynucleotide is
CC used for modulating, preferably increasing, the level of the synthase
CC in a plant cell. The plants are preferably monocots. The polynucleotide
CC is also used as a probe or primer in the detection quantitation or
CC isolation of gene transcripts. The probes are useful in detecting
CC deficiencies in the level of mRNA in screenings for desired transgenic
CC plant, for detecting mutations in the gene, for monitoring upregulation
CC of expression or changes in enzyme activity in screening assays of
CC compounds, for detection of any number of allelic variants of the gene,
CC or for use as molecular markers in plant breeding programs. The
CC isolated nucleic acids of the present invention can also be used for
CC recombinant expression of their encoded polypeptides or for use as
CC immunogens in the preparation and/or screening of antibodies. The
CC proteins can be employed in assays for enzyme agonists or antagonists
CC of enzyme function or for use of immunogens or antigens to obtain
CC antibodies specifically immunoreactive with a protein.
XX
SQ Sequence 3969 BP; 1022 A; 838 C; 1018 G; 1091 T; 0 other;

Query Match 43.3%; Score 1489.8; DB 21; Length 3969;
Best Local Similarity 67.3%; Pred. No. 0;
Matches 2189; Conservative 0; Mismatches 1032; Indels 30; Gaps 5;

```
Qy          9 GGTTAATCGCCGGTTCTCACAACAGGAATGAGTTTGTCCTCATTAATGCCGATGAGAATG 68
```

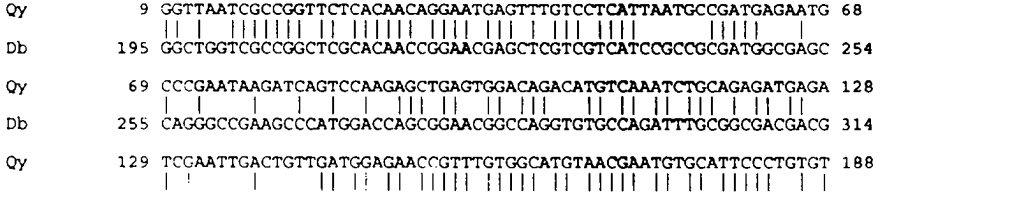


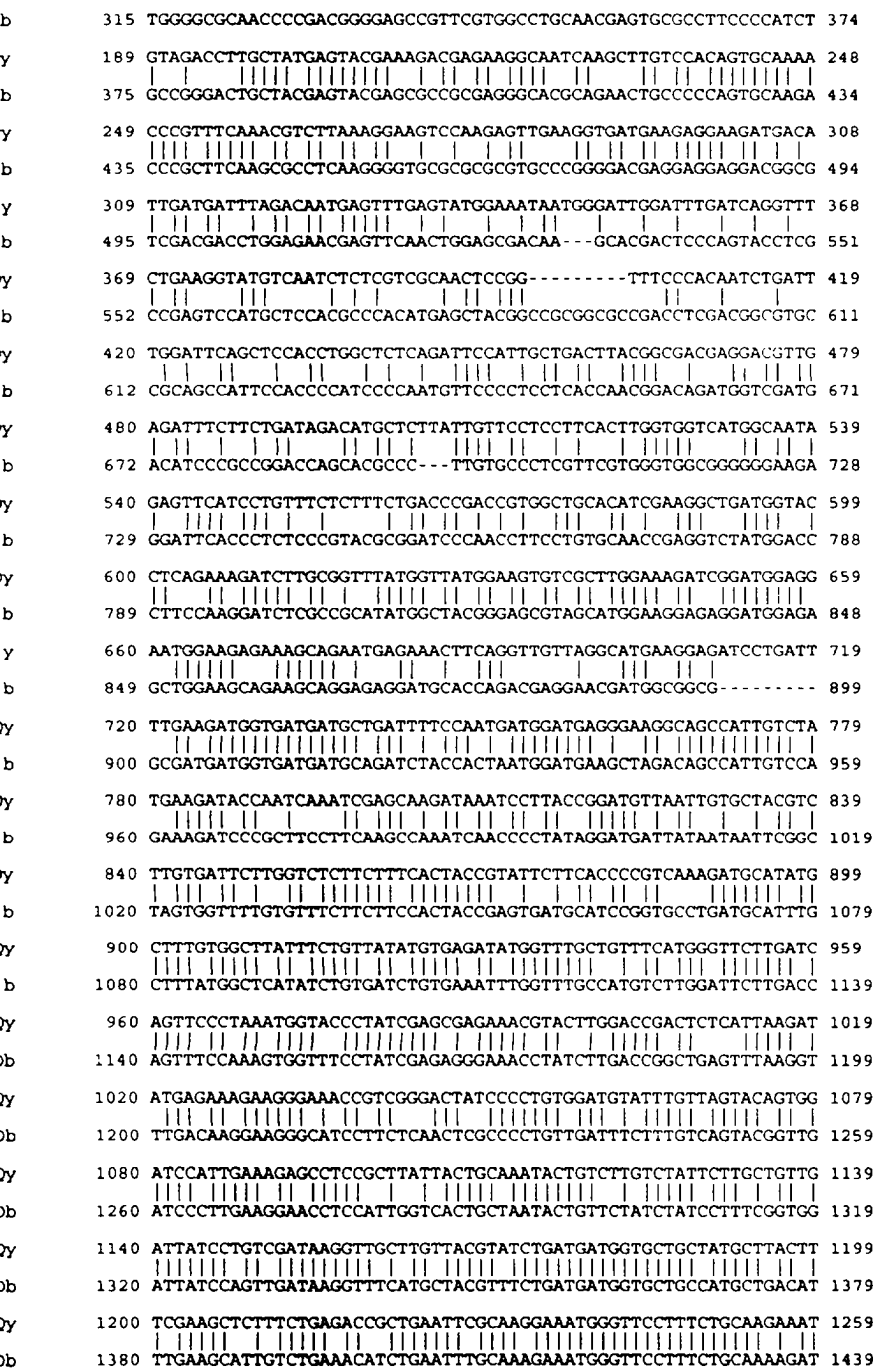


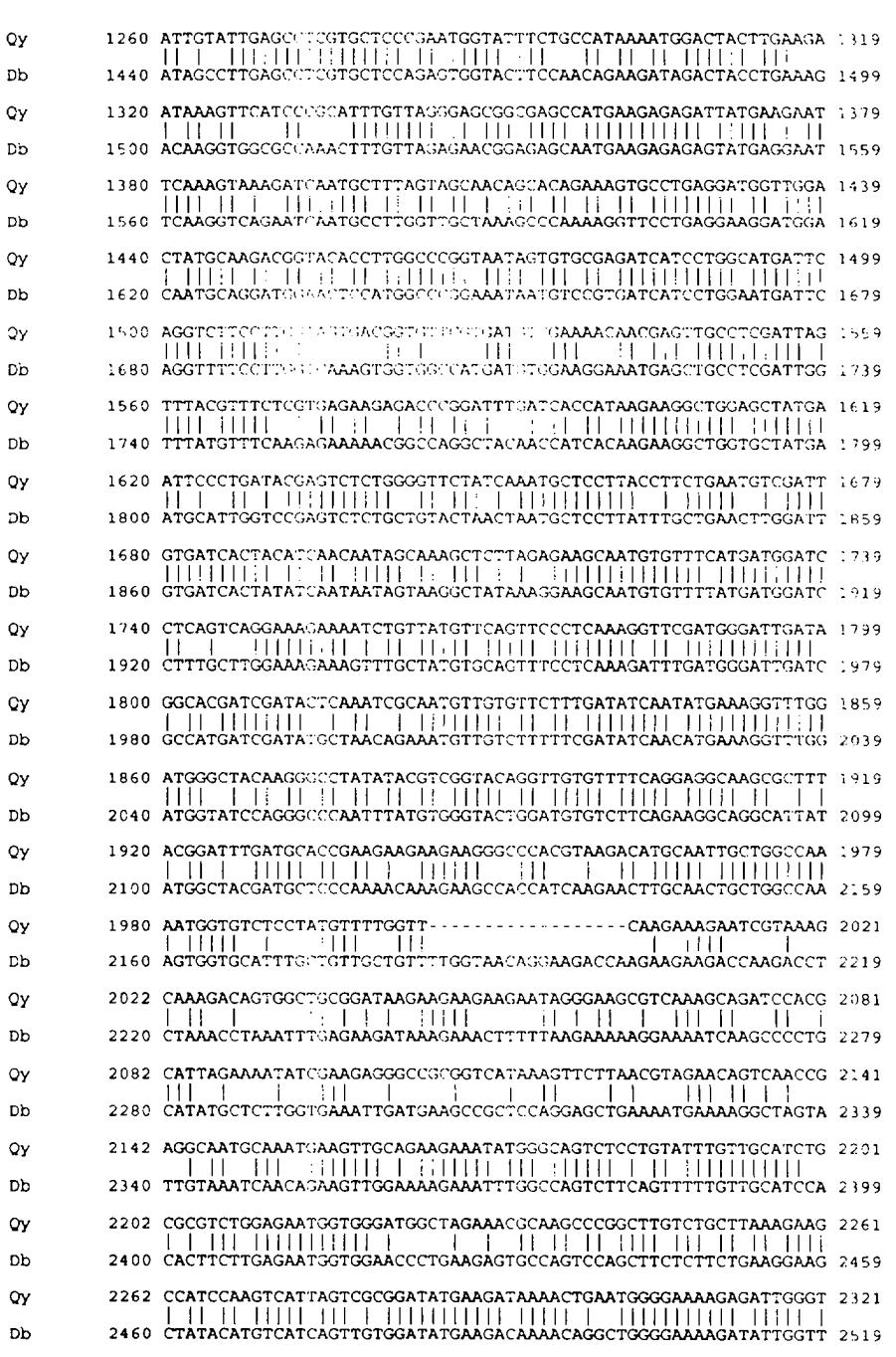


RESULT 8
AAZ58263
ID AAZ58263 standard; cDNA; 3776 BP.
XX
AC AAZ58263;
XX
DT 08-MAY-2000 (first entry)
XX
DE Corn cellulose synthase cDNA.
XX
KW Corn; maize; cellulose synthase; transgenic plant; ss.
XX
OS Zea mays.
XX
FH Key Location/Qualifiers
FT CDS 2..3448
FT /*tag= a
FT /partial
XX
PN WO200004166-A2.
XX
PD 27-JAN-2000.
XX
PF 13-JUL-1999; 99WO-US15871.
XX
PR 14-JUL-1998; 98US-0092844.
XX
PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX
PI Allen SM, Fader GM, Falco SC, Kinney AJ, Lightner JE, Miao G;
PI Rafalski JA, Thorpe CJ;
XX
DR WPI; 2000-182431/16.
DR P-PSDB; AAY58832.
XX
PT Nucleic acid fragments encoding cellulose biosynthetic enzyme useful as
PT probes for isolating cDNAs and genes encoding homologous proteins, for
PT producing transgenic plants -
XX
PS Claim 9; Page 37-38; 93pp; English.
XX
CC The present sequence is that of a contig of cDNA clones and encodes
CC a portion (see AAY58832) of corn cellulose synthase (CS). The cDNA
CC clones were isolated from corn cob, developing tassel, seedling,
CC shoot culture and leaf tissue cDNA libraries on the basis of
CC homology to Arabidopsis and cotton CS sequences. The invention
CC relates to isolated nucleic acid fragments encoding plant CS and to
CC CS polypeptides. It also relates to the construction of a chimeric
CC gene encoding all or a portion of the CS, in sense or antisense
CC orientation, where expression of the gene results in altered levels
CC of the CS in transformed host cells. The host cells can be used to
CC screen compounds for their ability to inhibit CS activity. CS
CC nucleic acids are also useful for producing transgenic plants having
CC altered levels of CS, and hence altered levels of fibre. CS may
CC also serve as a target for the development of novel herbicides.
XX
SQ Sequence 3776 BP; 931 A; 906 C; 999 G; 940 T; 0 other;

Query Match 43.2%; Score 1489; DB 21; Length 3776;
Best Local Similarity 67.5%; Pred. No. 0;
Matches 2203; Conservative 0; Mismatches 1015; Indels 45; Gaps 6;




b 315 TGGGGCGCAACCCCGACGGGGAGCCGTTCGTGGCCTGCAACGAGTGCGCCTTCCCCATCT 374
y 189 GTAGACCTTGCTATGAGTACGAAAGACGAGAAGGCAATCAAGCTTGTCCACAGTGCAAAA 248
b 375 GCCGGGACTGCTACGAGTACGAGCGCCGCGAGGGCACGCAGAACTGCCCCCAGTGCAAGA 434
y 249 CCCGTTTCAAACGTCTTAAAGGAAGTCCAAGAGTTGAAGGTGATGAAGAGGAAGATGACA 308
b 435 CCCGCTTCAAGCGCCTCAAGGGGTGCGCGCGCGTGCCCGGGGACGAGGAGGAGGACGGCG 494
y 309 TTGATGATTTAGACAATGAGTTTGAGTATGGAAATAATGGGATTGGATTTGATCAGGTTT 368
b 495 TCGACGACCTGGAGAACGAGTTCAACTGGAGCGACAA---GCACGACTCCCAGTACCTCG 551
y 369 CTGAAGGTATGTCAATCTCTCGTCGCAACTCCGG---------TTTCCCACAATCTGATT 419
b 552 CCGAGTCCATGCTCCACGCCCACATGAGCTACGGCCGCGGCGCCGACCTCGACGGCGTGC 611
y 420 TGGATTCAGCTCCACCTGGCTCTCAGATTCCATTGCTGACTTACGGCGACGAGGACGTTG 479
b 612 CGCAGCCATTCCACCCCATCCCCAATGTTCCCCTCCTCACCAACGGACAGATGGTCGATG 671
y 480 AGATTTCTTCTGATAGACATGCTCTTATTGTTCCTCCTTCACTTGGTGGTCATGGCAATA 539
b 672 ACATCCCGCCGGACCAGCACGCCC---TTGTGCCCTCGTTCGTGGGTGGCGGGGGGAAGA 728
y 540 GAGTTCATCCTGTTTCTCTTTCTGACCCGACCGTGGCTGCACATCGAAGGCTGATGGTAC 599
b 729 GGATTCACCCTCTCCCGTACGCGGATCCCAACCTTCCTGTGCAACCGAGGTCTATGGACC 788
y 600 CTCAGAAAGATCTTGCGGTTTATGGTTATGGAAGTGTCGCTTGGAAAGATCGGATGGAGG 659
b 789 CTTCCAAGGATCTCGCCGCATATGGCTACGGGAGCGTAGCATGGAAGGAGAGGATGGAGA 848
y 660 AATGGAAGAGAAAGCAGAATGAGAAACTTCAGGTTGTTAGGCATGAAGGAGATCCTGATT 719
b 849 GCTGGAAGCAGAAGCAGGAGAGGATGCACCAGACGAGGAACGATGGCGGCG--------- 899
y 720 TTGAAGATGGTGATGATGCTGATTTTCCAATGATGGATGAGGGAAGGCAGCCATTGTCTA 779
b 900 GCGATGATGGTGATGATGCAGATCTACCACTAATGGATGAAGCTAGACAGCCATTGTCCA 959
y 780 TGAAGATACCAATCAAATCGAGCAAGATAAATCCTTACCGGATGTTAATTGTGCTACGTC 839
b 960 GAAAGATCCCGCTTCCTTCAAGCCAAATCAACCCCTATAGGATGATTATAATAATTCGGC 1019
y 840 TTGTGATTCTTGGTCTCTTCTTTCACTACCGTATTCTTCACCCCGTCAAAGATGCATATG 899
b 1020 TAGTGGTTTTGTGTTTCTTCTTCCACTACCGAGTGATGCATCCGGTGCCTGATGCATTTG 1079
y 900 CTTTGTGGCTTATTTCTGTTATATGTGAGATATGGTTTGCTGTTTCATGGGTTCTTGATC 959
b 1080 CTTTATGGCTCATATCTGTGATCTGTGAAATTTGGTTTGCCATGTCTTGGATTCTTGACC 1139
y 960 AGTTCCCTAAATGGTACCCTATCGAGCGAGAAACGTACTTGGACCGACTCTCATTAAGAT 1019
b 1140 AGTTTCCAAAGTGGTTTCCTATCGAGAGGGAAACCTATCTTGACCGGCTGAGTTTAAGGT 1199
y 1020 ATGAGAAAGAAGGGAAACCGTCGGGACTATCCCCTGTGGATGTATTTGTTAGTACAGTGG 1079
b 1200 TTGACAAGGAAGGGCATCCTTCTCAACTCGCCCCTGTTGATTTCTTTGTCAGTACGGTTG 1259
y 1080 ATCCATTGAAAGAGCCTCCGCTTATTACTGCAAATACTGTCTTGTCTATTCTTGCTGTTG 1139
b 1260 ATCCCTTGAAGGAACCTCCATTGGTCACTGCTAATACTGTTCTATCTATCCTTTCGGTGG 1319
y 1140 ATTATCCTGTCGATAAGGTTGCTTGTTACGTATCTGATGATGGTGCTGCTATGCTTACTT 1199
b 1320 ATTATCCAGTTGATAAGGTTTCATGCTACGTTTCTGATGATGGTGCTGCCATGCTGACAT 1379
y 1200 TCGAAGCTCTTTCTGAGACCGCTGAATTCGCAAGGAAATGGGTTCCTTTCTGCAAGAAAT 1259
b 1380 TTGAAGCATTGTCTGAAACATCTGAATTTGCAAAGAAATGGGTTCCTTTCTGCAAAGAT 1439


Qy 1260 ATTGTATTGAGCCTGGTGCTCCCGAATGGTATTTCTGCCATAAAATGGACTACTTGAAGA 1319
Db 1440 ATAGCCTTGAGCCTCGTGCTCCAGAGTGGTACTTCCAACAGAAGATAGACTACCTGAAAG 1499
Qy 1320 ATAAAGTTCATCCGGCATTTGTTAGGGAGCGGCGAGCCATGAAGAGAGATTATGAAGAAT 1379
Db 1500 ACAAGGTGGCGCCAAACTTTGTTAGAGAACGGAGAGCAATGAAGAGAGAGTATGAGGAAT 1559
Qy 1380 TCAAAGTAAAGATCAATGCTTTAGTAGCAACAGCACAGAAAGTGCCTGAGGATGGTTGGA 1439
Db 1560 TCAAGGTCAGAATCAATGCCTTGGTTGCTAAAGCCCAAAAGGTTCCTGAGGAAGGATGGA 1619
Qy 1440 CTATGCAAGACGGTACACCTTGGCCCGGTAATAGTGTGCGAGATCATCCTGGCATGATTC 1499
Db 1620 CAATGCAGGATGGAACTTCATGGCCCGGAAATAATGTCCGTGATCATCCTGGAATGATTC 1679
Qy 1500 AGGTCTTCCTCTTTTTAATGACGGTGTTTTCGATTTTGAAAACAACGAGTTGCCTCGATTAG 1559
Db 1680 AGGTTTTCCTTTTTTAAAGTGGTGGCCATGATTGTGGAAGGAAATGAGCTGCCTCGATTGG 1739
Qy 1560 TTTACGTTTCTCGTGAGAAGACCCGGATTTGATCACCATAAGAAGGCTGGAGCTATGA 1619
Db 1740 TTTATGTTTCAAGAGAAAAACGGCCAGGCTACAACCATCACAAGAAGGCTGGTGCTATGA 1799
Qy 1620 ATTCCCTGATACGAGTCTCTGGGGTTCTATCAAATGCTCCTTACCTTCTGAATGTCGATT 1679
Db 1800 ATGCATTGGTCCGAGTCTCTGCTGTACTAACTAATGCTCCTTATTTGCTGAACTTGGATT 1859
Qy 1680 GTGATCACTACATCAACAATAGCAAAGCTCTTAGAGAAGCAATGTGTTTCATGATGGATC 1739
Db 1860 GTGATCACTATATCAATAATAGTAAGGCTATAAAGGAAGCAATGTGTTTTATGATGGATC 1919
Qy 1740 CTCAGTCAGGAAAGAAAATCTGTTATGTTCAGTTCCCTCAAAGGTTCGATGGGATTGATA 1799
Db 1920 CTTTGCTTGGAAAGAAAGTTTGCTATGTGCAGTTTCCTCAAAGATTTGATGGGATTGATC 1979
Qy 1800 GGCACGATCGATACTCAAATCGCAATGTTGTGTTCTTTGATATCAATATGAAAGGTTTGG 1859
Db 1980 GCCATGATCGATATGCTAACAGAAATGTTGTCTTTTTCGATATCAACATGAAAGGTTTGG 2039
Qy 1860 ATGGGCTACAAGGGCCTATATACGTCGGTACAGGTTGTGTTTTCAGGAGGCAAGCGCTTT 1919
Db 2040 ATGGTATCCAGGGCCCAATTTATGTGGGTACTGGATGTGTCTTCAGAAGGCAGGCATTAT 2099
Qy 1920 ACGGATTTGATGCACCGAAGAAGAAGAAGGGCCCACGTAAGACATGCAATTGCTGGCCAA 1979
Db 2100 ATGGCTACGATGCTCCCAAAACAAAGAAGCCACCATCAAGAACTTGCAACTGCTGGCCAA 2159
Qy 1980 AATGGTGTCTCCTATGTTTTGGTT----------CAAGAAAGAATCGTAAAG 2021
Db 2160 AGTGGTGCATTTGCTGTTGCTGTTTTGGTAACAGGAAGACCAAGAAGAAGACCAAGACCT 2219
Qy 2022 CAAAGACAGTGGCTGCGGATAAGAAGAAGAAGAATAGGGAAGCGTCAAAGCAGATCCACG 2081
Db 2220 CTAAACCTAAATTTGAGAAGATAAAGAAACTTTTAAGAAAAAGGAAAATCAAGCCCCTG 2279
Qy 2082 CATTAGAAAATATCGAAGAGGGCCGCGGTCATAAAGTTCTTAACGTAGAACAGTCAACCG 2141
Db 2280 CATATGCTCTTGGTGAAATTGATGAAGCCGCTCCAGGAGCTGAAAATGAAAAGGCTAGTA 2339
Qy 2142 AGGCAATGCAAATGAAGTTGCAGAAGAAATATGGGCAGTCTCCTGTATTTGTTGCATCTG 2201
Db 2340 TTGTAAATCAACAGAAGTTGGAAAAGAAATTTGGCCAGTCTTCAGTTTTTGTTGCATCCA 2399
Qy 2202 CGCGTCTGGAGAATGGTGGGATGGCTAGAAACGCAAGCCCGGCTTGTCTGCTTAAAGAAG 2261
Db 2400 CACTTCTTGAGAATGGTGGAACCCTGAAGAGTGCCAGTCCAGCTTCTCTTCTGAAGGAAG 2459
Qy 2262 CCATCCAAGTCATTAGTCGCGGATATGAAGATAAAACTGAATGGGGAAAAGAGATTGGGT 2321
Db 2460 CTATACATGTCATCAGTTGTGGATATGAAGACAAAACAGGCTGGGGAAAAGATATTGGTT 2519

```
Qy       2322 GGATCTATGGTTCTGTTACCGAAGATATTCTTACGGGTTCTAAGATGCATTCTCATGGTT 2381
Db       2520 GGATTTATGGATCAGTCACAGAAGATATTCTTACTGGGTTTAAGATGCACTGCCATGGTT 2579

Qy       2382 GGAGACATGTTTATTGTACACCAAAGTTAGCGGCTTTCAAAGGATCAGCTCCAATCAATC 2441
Db       2580 GGCGGTCAATTTACTGCATACCTAAACGGGCCGCCTTCAAAGGTTCCGCACCTCTCAATC 2639

Qy       2442 TTTCGGATCGTCTCCATCAAGTTCTTCGATGGGCGCTTGGGTCGGTTGAGATTTTCTTGA 2501
Db       2640 TTTCCGATCGTCTTCACCAGGTTCTTCGGTGGGCTCTTGGTTCAATTGAAATTTTCTTCA 2699

Qy       2502 GTAGGCATTGTCCTATTTGGTATGGTTATGGAGGTGGGTTGAAATGGCTTGAGCGGTTGT 2561
Db       2700 GCAACCACTGCCCTCTCTGGTATGGGTATGGTGGTGGACTAAAGTTCCTGGAAAGGTTTT 2759

Qy       2562 CCTACATTAACTCTGTGGTTTACCCGTGGACCTCTCTACCGCTCATCGTTTACTGTTCTC 2621
Db       2760 CGTACATTAACTCCATCGTATACCCTTGGACATCTATCCCGCTCTTGGCCTATTGCACAT 2819

Qy       2622 TCCCTGCCATCTGTCTTCTCACTGGAAAATTCATCGTTCCCGAGATTAGCAACTATGCGA 2681
Db       2820 TGCCTGCCATCTGCTTGCTGACAGGGAAATTTATCACGCCAGAGCTTAACAATGTTGCCA 2879

Qy       2682 GTATCCTCTTCATGGCGCTCTTCTCGTCGATTGCAATAACGGGTATTCTCGAGATGCAAT 2741
Db       2880 GCCTCTGGTTCATGTCACTTTTCATCTGCATTTTTGCTACGAGCATCCTGGAAATGAGAT 2939

Qy       2742 GGGGCAAAGTTGGGATCGATGATTGGTGGAGAAACGAACAGTTTTGGGTCATTGGAGGTG 2801
Db       2940 GGAGTGGTGTAGGCATCGATGACTGGTGGAGAAACGAGCAGTTTTGGGTCATTGGAGGCG 2999

Qy       2802 TTTCTGCGCATCTGTTTGCTCTCTTCCAAGGTCTCCTCAAGGTTCTTGCTGGTGTCGACA 2861
Db       3000 TGTCTTCACATCTCTTTGCTGTGTTCCAGGGACTCCTCAAGGTCATAGCTGGTGTAGACA 3059

Qy       2862 CTAACTTCACAGTCACATCAAAAGCAGCTGATGATGGAGAGTTCTCTGACCTTTACCTCT 2921
Db       3060 CGAGCTTCACTGTGACATCCAAGGGCGGAGACGACGAGGAGTTCTCAGAGCTGTACACAT 3119

Qy       2922 TCAAATGGACTTCACTTCTCATCCCTCCAATGACTCTACTCATCATAAACGTCATTGGAG 2981
Db       3120 TCAAATGGACGACCCTTCTGATACCTCCGACAACCCTGCTCCTACTGAACTTCATTGGAG 3179

Qy       2982 TCATAGTCGGAGTCTCTGATGCCATCAGCAATGGATACGACTCGTGGGGACCGCTTTTCG 3041
Db       3180 TGGTAGCTGGCATCTCCAATGCGATCAACAACGGATATGAATCATGGGGCCCCCTGTTCG 3239

Qy       3042 GAAGACTGTTCTTTGCACTTTGGGTCATCATTCATCTTTACCGTTCCTTAAAGGTTTGC 3101
Db       3240 GGAAGCTCTTCTTTGCATTTTGGGTGATCGTCCATCTTTACGTTCCTCAAGGGTCTGG 3299

Qy       3102 TTGGGAAACAAGATAGAATGCCAACCATTATTGTCGTCTGGTCATCCTCCTGGCCTCGA 3161
Db       3300 TTGGGAGGCAGAACAGGACGCCAACGATTGTCATTGTCTGGTCCATCCTCCTGGCTTCGA 3359

Qy       3162 TTCTTACACTTCTTTGGGTCCGGGTTAATCCGTTTGTGGCGA---AAGGCGGTCCTATTC 3218
Db       3360 TCTTCTCGCTGCTTTGGGTCCGGATCGACCCGTTCCTTGCGAAGGATGATGGTCCCCTGT 3419

Qy       3219 TCGAGATCTGTGGTTTAGACTGC 3241
Db       3420 TGGAGGAGTGTGGTCTGGATTGC 3442

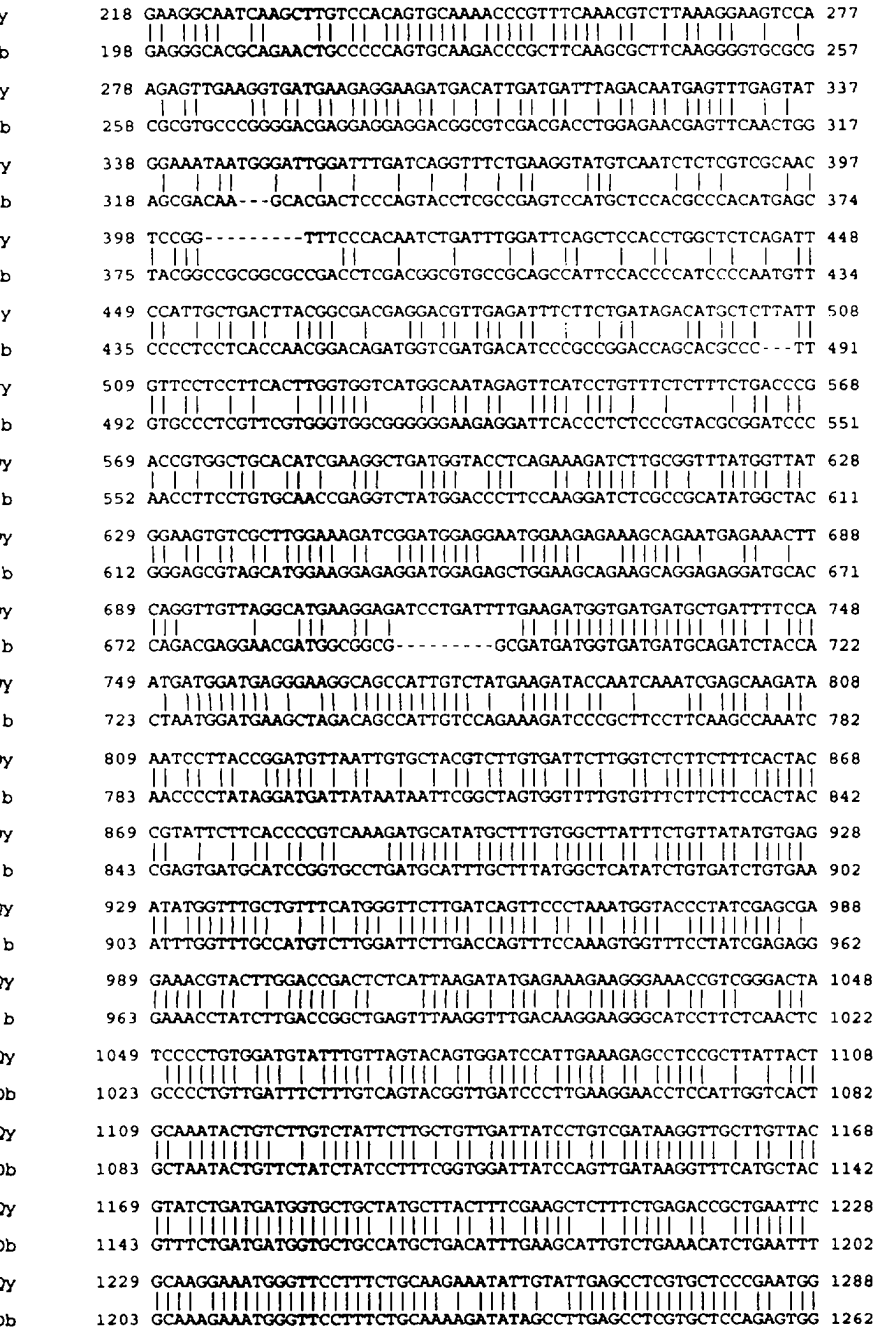

RESULT 9
AAZ99491
ID   AAZ99491 standard; DNA; 3568 BP.
XX
AC   AAZ99491;
XX
DT   03-JUL-2000  (first entry)
XX
DE   DNA encoding a maize cellulose synthase.
XX
KW   Maize; cellulose synthase; stalk quality; stand; silage; cellulose;
KW   transgenic plant; plant breeding marker; ss.
XX
OS   Zea mays.
XX
FH   Key             Location/Qualifiers
FT   CDS             63..3236
FT                   /*tag=  a
FT                   /product= "cellulose synthase"
FT                   /note= "no termination codon given"
XX
PN   WO200009706-A2.
XX
PD   24-FEB-2000.
XX
PF   16-AUG-1999;   99WO-US18760.
XX
PR   17-AUG-1998;   98US-0096822.
XX
PA   (PION-) PIONEER HI-BRED INT INC.
XX
PI   Dhugga KS,  Helentjaris TG,  Bowen BA,  Wang X;
XX
DR   WPI; 2000-224343/19.
DR   P-PSDB; AAY84107.
XX
PT   New genes which encode maize cellulose synthase polypeptides in plants
PT   useful for modulating the expression of cellulose synthase in plants
PT   and to produce transgenic plants expressing the novel protein  -
XX
PS   Claim 1; Page 82-87; 119pp; English.
XX
CC   The present sequence encodes a maize cellulose synthase polypeptide.
CC   The cellulose synthase can be used for the improvement of stalk quality
CC   for improved stand or silage. It also provides an increased concentration
CC   of cellulose in the pericarp, hardening the kernel and improving its
CC   handling ability. The sequences are used to produce transgenic plants
CC   and seeds expressing the cellulose synthase. The polynucleotide is
CC   used for modulating, preferably increasing, the level of the synthase
CC   in a plant cell. The plants are preferably monocots. The polynucleotide
CC   is also used as a probe or primer in the detection quantitation or
CC   isolation of gene transcripts. The probes are useful in detecting
CC   deficiencies in the level of mRNA in screenings for desired transgenic
CC   plant, for detecting mutations in the gene, for monitoring upregulation
CC   of expression or changes in enzyme activity in screening assays of
CC   compounds, for detection of any number of allelic variants of the gene,
CC   or for use as molecular markers in plant breeding programs. The
CC   isolated nucleic acids of the present invention can also be used for
CC   recombinant expression of their encoded polypeptides or for use as
CC   immunogens in the preparation and/or screening of antibodies. The
CC   proteins can be employed in assays for enzyme agonists or antagonists
CC   of enzyme function or for use of immunogens or antigens to obtain
CC   antibodies specifically immunoreactive with a protein.
XX
SQ   Sequence 3568 BP; 879 A; 825 C; 941 G; 922 T; 1 other;

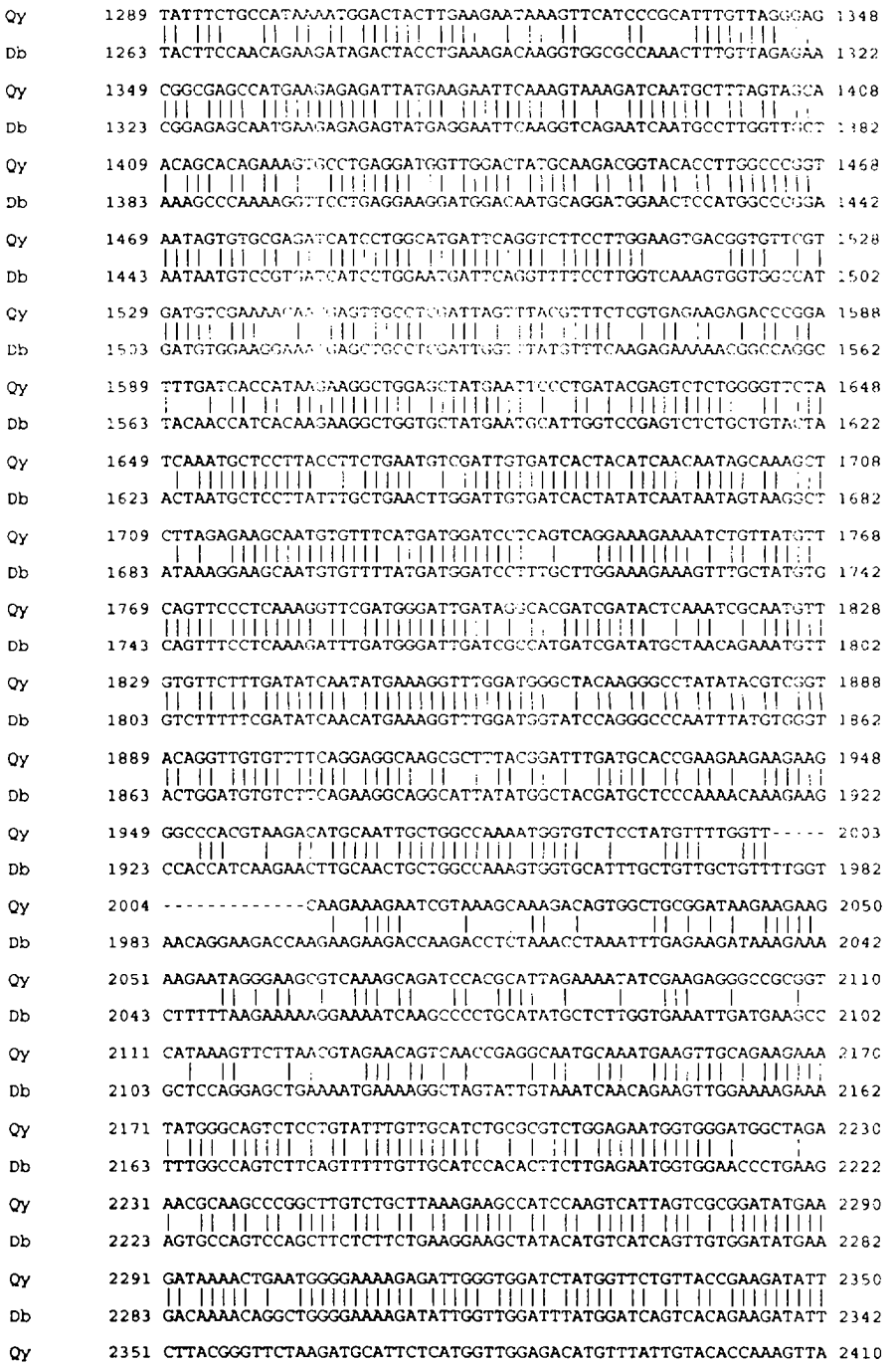
  Query Match           42.5%;  Score 1464.8;  DB 21;  Length 3568;
  Best Local Similarity   67.3%;  Pred. No. 0;
  Matches 2177;  Conservative    0;  Mismatches 1012;  Indels   45;  Gaps    6;

Qy         38 GAGTTTGTCCTCATTAATGCCGATGAGAATGCCCGAATAAGATCAGTCCAAGAGCTGAGT 97
Db         18 GAGCTCGTCGTCATCCGCCGCGATGGCGAGCCAGGGCCGAAGCCCATGGACCAGCGGAAC 77

Qy         98 GGACAGACATGTCAAATCTGCAGAGATGAGATCGAATTGACTGTTGATGGAGAACCGTTT 157
Db         78 GGCCAGGTGTGCCAGATTTGCGGCGACGACGTGGGGCGCAACCCCGACGGGGAGCCTTTC 137

Qy        158 GTGGCATGTAACGAATGTGCATTCCCTGTGTGTAGACCTTGCTATGAGTACGAAAGACGA 217
Db        138 GTGGCCTGCAACGAGTGCGCCTTCCCCATCTGCCGGGACTGCTACGAGTACGAGCGCCGC 197
```


Qy 218 GAAGGCAATCAAGCTTGTCCACAGTGCAAAACCCGTTTCAAACGTCTTAAAGGAAGTCCA 277
Db 198 GAGGGCACGCAGAACTGCCCCCAGTGCAAGACCCGCTTCAAGCGCTTCAAGGGGTGCGCG 257
Qy 278 AGAGTTGAAGGTGATGAAGAGGAAGATGACATTGATGATTTAGACAATGAGTTTGAGTAT 337
Db 258 CGCGTGCCCGGGGACGAGGAGGAGGACGGCGTCGACGACCTGGAGAACGAGTTCAACTGG 317
Qy 338 GGAAATAATGGGATTGGATTTGATCAGGTTTCTGAAGGTATGTCAATCTCTCGTCGCAAC 397
Db 318 AGCGACAA---GCACGACTCCCAGTACCTCGCCGAGTCCATGCTCCACGCCCACATGAGC 374
Qy 398 TCCGG---------TTTCCCACAATCTGATTTGGATTCAGCTCCACCTGGCTCTCAGATT 448
Db 375 TACGGCCGCGGCGCCGACCTCGACGGCGTGCCGCAGCCATTCCACCCCATCCCCAATGTT 434
Qy 449 CCATTGCTGACTTACGGCGACGAGGACGTTGAGATTTCTTCTGATAGACATGCTCTTATT 508
Db 435 CCCCTCCTCACCAACGGACAGATGGTCGATGACATCCCGCCGGACCAGCACGCCC---TT 491
Qy 509 GTTCCTCCTTCACTTGGTGGTCATGGCAATAGAGTTCATCCTGTTTCTCTTTCTGACCCG 568
Db 492 GTGCCCTCGTTCGTGGGTGGCGGGGGGAAGAGGATTCACCCTCTCCCGTACGCGGATCCC 551
Qy 569 ACCGTGGCTGCACATCGAAGGCTGATGGTACCTCAGAAAGATCTTGCGGTTTATGGTTAT 628
Db 552 AACCTTCCTGTGCAACCGAGGTCTATGGACCCTTCCAAGGATCTCGCCGCATATGGCTAC 611
Qy 629 GGAAGTGTCGCTTGGAAAGATCGGATGGAGGAATGGAAGAGAAAGCAGAATGAGAAACTT 688
Db 612 GGGAGCGTAGCATGGAAGGAGAGGATGGAGAGCTGGAAGCAGAAGCAGGAGAGGATGCAC 671
Qy 689 CAGGTTGTTAGGCATGAAGGAGATCCTGATTTTGAAGATGGTGATGATGCTGATTTTCCA 748
Db 672 CAGACGAGGAACGATGGCGGCG---------GCGATGATGGTGATGATGCAGATCTACCA 722
Qy 749 ATGATGGATGAGGGAAGGCAGCCATTGTCTATGAAGATACCAATCAAATCGAGCAAGATA 808
Db 723 CTAATGGATGAAGCTAGACAGCCATTGTCCAGAAAGATCCCGCTTCCTTCAAGCCAAATC 782
Qy 809 AATCCTTACCGGATGTTAATTGTGCTACGTCTTGTGATTCTTGGTCTCTTCTTTCACTAC 868
Db 783 AACCCCTATAGGATGATTATAATAATTCGGCTAGTGGTTTTGTGTTTCTTCTTCCACTAC 842
Qy 869 CGTATTCTTCACCCCGTCAAAGATGCATATGCTTTGTGGCTTATTTCTGTTATATGTGAG 928
Db 843 CGAGTGATGCATCCGGTGCCTGATGCATTTGCTTTATGGCTCATATCTGTGATCTGTGAA 902
Qy 929 ATATGGTTTGCTGTTTCATGGGTTCTTGATCAGTTCCCTAAATGGTACCCTATCGAGCGA 988
Db 903 ATTTGGTTTGCCATGTCTTGGATTCTTGACCAGTTTCCAAAGTGGTTTCCTATCGAGAGG 962
Qy 989 GAAACGTACTTGGACCGACTCTCATTAAGATATGAGAAAGAAGGGAAACCGTCGGGACTA 1048
Db 963 GAAACCTATCTTGACCGGCTGAGTTTAAGGTTTGACAAGGAAGGGCATCCTTCTCAACTC 1022
Qy 1049 TCCCCTGTGGATGTATTTGTTAGTACAGTGGATCCATTGAAAGAGCCTCCGCTTATTACT 1108
Db 1023 GCCCCTGTTGATTTCTTTGTCAGTACGGTTGATCCCTTGAAGGAACCTCCATTGGTCACT 1082
Qy 1109 GCAAATACTGTCTTGTCTATTCTTGCTGTTGATTATCCTGTCGATAAGGTTGCTTGTTAC 1168
Db 1083 GCTAATACTGTTCTATCTATCCTTTCGGTGGATTATCCAGTTGATAAGGTTTCATGCTAC 1142
Qy 1169 GTATCTGATGATGGTGCTGCTATGCTTACTTTCGAAGCTCTTTCTGAGACCGCTGAATTC 1228
Db 1143 GTTTCTGATGATGGTGCTGCCATGCTGACATTTGAAGCATTGTCTGAAACATCTGAATTT 1202
Qy 1229 GCAAGGAAATGGGTTCCTTTCTGCAAGAAATATTGTATTGAGCCTCGTGCTCCCGAATGG 1288
Db 1203 GCAAAGAAATGGGTTCCTTTCTGCAAAAGATATAGCCTTGAGCCTCGTGCTCCAGAGTGG 1262


Qy 1289 TATTTCTGCCATAAAATGGACTACTTGAAGAATAAAGTTCATCCCGCATTTGTTAGGGAG 1348
Db 1263 TACTTCCAACAGAAGATAGACTACCTGAAAGACAAGGTGGCGCCAAACTTTGTTAGAGAA 1322
Qy 1349 CGGCGAGCCATGAAGAGAGATTATGAAGAATTCAAAGTAAAGATCAATGCTTTAGTAGCA 1408
Db 1323 CGGAGAGCAATGAAGAGAGAGTATGAGGAATTCAAGGTCAGAATCAATGCCTTGGTTGCT 1382
Qy 1409 ACAGCACAGAAAGTGCCTGAGGATGGTTGGACTATGCAAGACGGTACACCTTGGCCCGGT 1468
Db 1383 AAAGCCCAAAAGGTTCCTGAGGAAGGATGGACAATGCAGGATGGAACTCCATGGCCCGGA 1442
Qy 1469 AATAGTGTGCGAGATCATCCTGGCATGATTCAGGTCTTCCTTGGAAGTGACGGTGTTCGT 1528
Db 1443 AATAATGTCCGTGATCATCCTGGAATGATTCAGGTTTTCCTTGGTCAAAGTGGTGGCCAT 1502
Qy 1529 GATGTCGAAAACAAGAGTTGCCTCGATTAGTTTACGTTTCTCGTGAGAAGAGACCCGGA 1588
Db 1503 GATGTGGAAGGAAAGAGCTGCCTCGATTGGTGTATGTTTCAAGAGAAAAACGGCCAGGC 1562
Qy 1589 TTTGATCACCATAAGAAGGCTGGAGCTATGAATTCCCTGATACGAGTCTCTGGGGTTCTA 1648
Db 1563 TACAACCATCACAAGAAGGCTGGTGCTATGAATGCATTGGTCCGAGTCTCTGCTGTACTA 1622
Qy 1649 TCAAATGCTCCTTACCTTCTGAATGTCGATTGTGATCACTACATCAACAATAGCAAAGCT 1708
Db 1623 ACTAATGCTCCTTATTTGCTGAACTTGGATTGTGATCACTATATCAATAATAGTAAGGCT 1682
Qy 1709 CTTAGAGAAGCAATGTGTTTCATGATGGATCCTCAGTCAGGAAAGAAAATCTGTTATGTT 1768
Db 1683 ATAAAGGAAGCAATGTGTTTTATGATGGATCCTTTGCTTGGAAAGAAAGTTTGCTATGTG 1742
Qy 1769 CAGTTCCCTCAAAGGTTCGATGGGATTGATAGGCACGATCGATACTCAAATCGCAATGTT 1828
Db 1743 CAGTTTCCTCAAAGATTTGATGGGATTGATCGCCATGATCGATATGCTAACAGAAATGTT 1802
Qy 1829 GTGTTCTTTGATATCAATATGAAAGGTTTGGATGGGCTACAAGGGCCTATATACGTCGGT 1888
Db 1803 GTCTTTTTCGATATCAACATGAAAGGTTTGGATGGTATCCAGGGCCCAATTTATGTGGGT 1862
Qy 1889 ACAGGTTGTGTTTTCAGGAGGCAAGCGCTTTACGGATTTGATGCACCGAAGAAGAAGAAG 1948
Db 1863 ACTGGATGTGTCTTCAGAAGGCAGGCATTATATGGCTACGATGCTCCCAAAACAAAGAAG 1922
Qy 1949 GGCCCACGTAAGACATGCAATTGCTGGCCAAAATGGTGTCTCCTATGTTTTGGTT----- 2003
Db 1923 CCACCATCAAGAACTTGCAACTGCTGGCCAAAGTGGTGCATTTGCTGTTGCTGTTTTGGT 1982
Qy 2004 -------------CAAGAAAGAATCGTAAAGCAAAGACAGTGGCTGCGGATAAGAAGAAG 2050
Db 1983 AACAGGAAGACCAAGAAGAAGACCAAGACCTCTAAACCTAAATTTGAGAAGATAAAGAAA 2042
Qy 2051 AAGAATAGGGAAGCGTCAAAGCAGATCCACGCATTAGAAAATATCGAAGAGGGCCGCGGT 2110
Db 2043 CTTTTTAAGAAAAAGGAAAATCAAGCCCCTGCATATGCTCTTGGTGAAATTGATGAAGCC 2102
Qy 2111 CATAAAGTTCTTAACGTAGAACAGTCAACCGAGGCAATGCAAATGAAGTTGCAGAAGAAA 2170
Db 2103 GCTCCAGGAGCTGAAAATGAAAAGGCTAGTATTGTAAATCAACAGAAGTTGGAAAAGAAA 2162
Qy 2171 TATGGGCAGTCTCCTGTATTTGTTGCATCTGCGCGTCTGGAGAATGGTGGGATGGCTAGA 2230
Db 2163 TTTGGCCAGTCTTCAGTTTTTGTTGCATCCACACTTCTTGAGAATGGTGGAACCCTGAAG 2222
Qy 2231 AACGCAAGCCCGGCTTGTCTGCTTAAAGAAGCCATCCAAGTCATTAGTCGCGGATATGAA 2290
Db 2223 AGTGCCAGTCCAGCTTCTCTTCTGAAGGAAGCTATACATGTCATCAGTTGTGGATATGAA 2282
Qy 2291 GATAAAACTGAATGGGGAAAAGAGATTGGGTGGATCTATGGTTCTGTTACCGAAGATATT 2350
Db 2283 GACAAAACAGGCTGGGGAAAAGATATTGGTTGGATTTATGGATCAGTCACAGAAGATATT 2342
Qy 2351 CTTACGGGTTCTAAGATGCATTCTCATGGTTGGAGACATGTTTATTGTACACCAAAGTTA 2410

```
Db       2343 CTTACTGGGTTTAAGATGCACTGCCATGGTTGGCGGTCAATTTACTGCATACCTAAACGG 2402

Qy       2411 GCGGCTTTCAAAGGATCAGCTCCAATCAATCTTTCGGATCGTCTCCATCAAGTTCTTCGA 2470

Db       2403 GCCGCCTTCAAAGGTTCCGCACCTCTCAATCTTTCCGATCGTTTTCACCAGGTTCTTCGG 2462

Qy       2471 TGGGCGCTTGGGTCGGTTGAGATTTTCTTGAGTAGGCATTGTCCTATTTGGTATGGTTAT 2530

Db       2463 TGGGCTCTTGGTTCAATTGAAATTTTGTTCAGCAACCACTGCCCTCTCTGGTATGGGTAT 2522

Qy       2531 GGAGGTGGGTTGAAATGGCTTGAGCGGTTGTCCTACATTAACTCTGTGGTTTACCCGTGG 2590

Db       2523 GGTGGTGGACTAAAGTTCCTGGAAAGGTTTTCGTACATTAACTCCATCGTATACCCTTGG 2582

Qy       2591 ACCTCTCTACCGCTCATCGTTTACTGTTCTCTCCCTGCCATCTGTCTTCTCACTGGAAAA 2650

Db       2583 ACATCTATCCCGCTCTTGGCCTATTGCACATTGCCTGCCATCTGCTTGCTGACAGGGAAA 2642

Qy       2651 TTCATCGTTCCCGAGATTAGCAACTATGCGAGTATCCTCTTCATGGCGCTCTTCTCGTCG 2710

Db       2643 TTTATCACGCCAGAGCTTAACAATGTTGCCAGCCTCTGGTTCATGTCACTTTTCATCTGC 2702

Qy       2711 ATTGCAATAACGGGTATTCTCGAGATGCAATGGGGCAAAGTTGGGATCGATGATTGGTGG 2770

Db       2703 ATTTTTGCTACGAGCATCCTGGAAATGAGATGGAGTGGTGTAGGCATCGATGACTGGTGG 2762

Qy       2771 AGAAACGAACAGTTTTGGGTCATTGGAGGTGTTTCTGCGCATCTGTTTGCTCTCTTCCAA 2830

Db       2763 AGAAACGAGCAGTTTTGGGTCATTGGAGGCGTGTCTTCACATCTCTTTGCTGTGTTCCAG 2822

Qy       2831 GGTCTCCTCAAGGTTCTTGCTGGTGTCGACACTAACTTCACAGTCACATCAAAAGCAGCT 2890

Db       2823 GGACTCCTCAAGGTCATAGCTGGTGTAGACACGAGCTTCACTGTGACATCCAAGGGCGGA 2882

Qy       2891 GATGATGGAGAGTTCTCTGACCTTTACCTCTTCAAATGGACTTCACTTCTCATCCCTCCA 2950

Db       2883 GACGACGAGGAGTTCTCAGAGCTGTACACATTCAAATGGACGACCCTTCTGATACCTCCG 2942

Qy       2951 ATGACTCTACTCATCATAAACGTCATTGGAGTCATAGTCGGAGTCTCTGATGCCATCAGC 3010

Db       2943 ACAACCCTGCTCCTACTGAACTTCATTGGAGTGGTAGCTGGCATCTCCAATGCGATCAAC 3002

Qy       3011 AATGGATACGACTCGTGGGGACCGCTTTTCGGAAGACTGTTCTTTGCACTTTGGGTCATC 3070

Db       3003 AACGGATATGAATCATGGGGCCCCCTGTTCGGGAAGCTCTTCTTTGCATTTTGGGTGATC 3062

Qy       3071 ATTCATCTTTACCCGTTCCTTAAAGGTTTGCTTGGGAAACAAGATAGAATGCCAACCATT 3130

Db       3063 GTCCATCTTTACCCGTTCCTCAAGGGTCTGGTTGGGAGGCAGAACAGGACGCCAACGATT 3122

Qy       3131 ATTGTCGTCTGGTCCATCCTCCTGGCCTGGATTCTTACACTCTTTGGGTCCGGGTTAAT 3190

Db       3123 GTCATTGTCTGGTCCATCCTCCTGGCTTCGATCTTCTCGCTGCTTTGGGTCCGGATCGAC 3182

Qy       3191 CCGTTTGTGGCGA---AAGGCGGTCCTATTCTCGAGATCTGTGGTTTAGACTGC 3241

Db       3183 CCGTTCCTTGCGAAGGATGATGGTCCCCTGTTGGAGGAGTGTGGTCTGGATTGC 3236


RESULT 10
AAZ58270
ID   AAZ58270 standard; cDNA; 2890 BP.
XX
AC   AAZ58270;
XX
DT   08-MAY-2000  (first entry)
XX
DE   Soybean cellulose synthase cDNA clone ss1.pk0036.c10.
XX
KW   Soybean; cellulose synthase; transgenic plant; ss.
XX
OS   Glycine max.
XX
FH   Key             Location/Qualifiers
FT   CDS             2..2383
FT                   /*tag=  a
FT                   /partial
XX
PN   WO200004166-A2.
XX
PD   27-JAN-2000.
XX
PF   13-JUL-1999;   99WO-US15871.
XX
PR   14-JUL-1998;   98US-0092844.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Allen SM,  Fader GM,  Falco SC,  Kinney AJ,  Lightner JE,  Miao G;
PI   Rafalski JA,  Thorpe CJ;
XX
DR   WPI; 2000-182431/16.
DR   P-PSDB; AAY58839.
XX
PT   Nucleic acid fragments encoding cellulose biosynthetic enzyme useful as
PT   probes for isolating cDNAs and genes encoding homologous proteins, for
PT   producing transgenic plants  -
XX
PS   Claim 3; Page 62-63; 93pp; English.
XX
CC   The present sequence is that of cDNA clone ss1.pk0036.c10 encoding
CC   a portion (see AAY58839) of soybean cellulose synthase (CS).  The
CC   cDNA clone was isolated from a cDNA library constructed from
CC   soybean seedling (5-10 day) tissue on the basis of homology to
CC   Arabidopsis and cotton CS sequences.  The invention relates to
CC   isolated nucleic acid fragments encoding plant CS and to CS
CC   polypeptides.  It also relates to the construction of a chimeric
CC   gene encoding all or a portion of the CS, in sense or antisense
CC   orientation, where expression of the gene results in altered levels
CC   of the CS in transformed host cells.  The host cells can be used to
CC   screen compounds for their ability to inhibit CS activity.  CS
CC   nucleic acids are also useful for producing transgenic plants having
CC   altered levels of CS, and hence altered levels of fibre.  CS may
CC   also serve as a target for the development of novel herbicides.
XX
SQ   Sequence 2890 BP; 808 A; 516 C; 695 G; 871 T; 0 other;

  Query Match             42.3%;  Score 1458.4;  DB 21;  Length 2890;
  Best Local Similarity   76.6%;  Pred. No. 0;
  Matches 1815;  Conservative    0;  Mismatches  546;  Indels    9;  Gaps    2;

Qy        875 CTTCACCCCGTCAAAGATGCATATGCTTTGTGGCTTATTTCTGTTATATGTGAGATATGG 934

Db          8 CTCCACCCAGTTAATGATGCATATGGCTTGTGGTTGACATCAGTCATCTGTGAAATATGG 67

Qy        935 TTTGCTGTTTCATGGGTTCTTGATCAGTTCCCTAAATGGTACCCTATCGAGCGAGAAACG 994

Db         68 TTTGCTGTATCATGGATAATGGATCAGTTTCCAAAATGGTACCCAATACAGCGAGAAACA 127

Qy        995 TACTTGGACCGACTCTCATTAAGATATGAGAAAGAAGGGAAACCGTCGGGACTATCCCCT 1054

Db        128 TACCTTGATCGTCTGTCACTCAGGTATGAAAAAGAAGGGAAGCCATCTGAGTTGTCCAGT 187

Qy       1055 GTGGATGTATTTGTTAGTACAGTGGATCCATTGAAAGAGCCTCCGCTTATTACTGCAAAT 1114

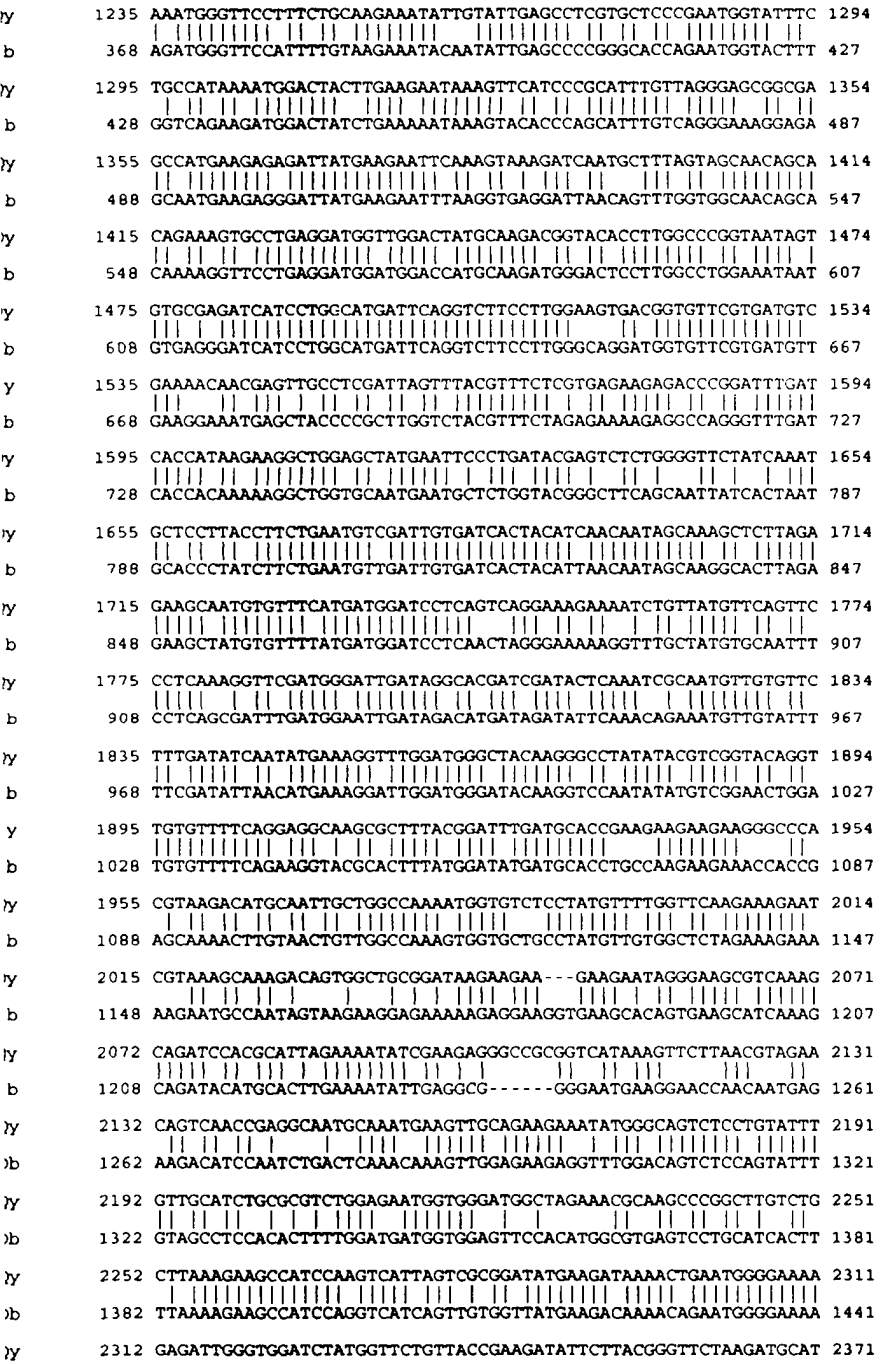
Db        188 GTAGACGTCTTTGTCAGTACTGTTGATCCCATGAAGGAACCTCCACTGATTACAGCAAAC 247

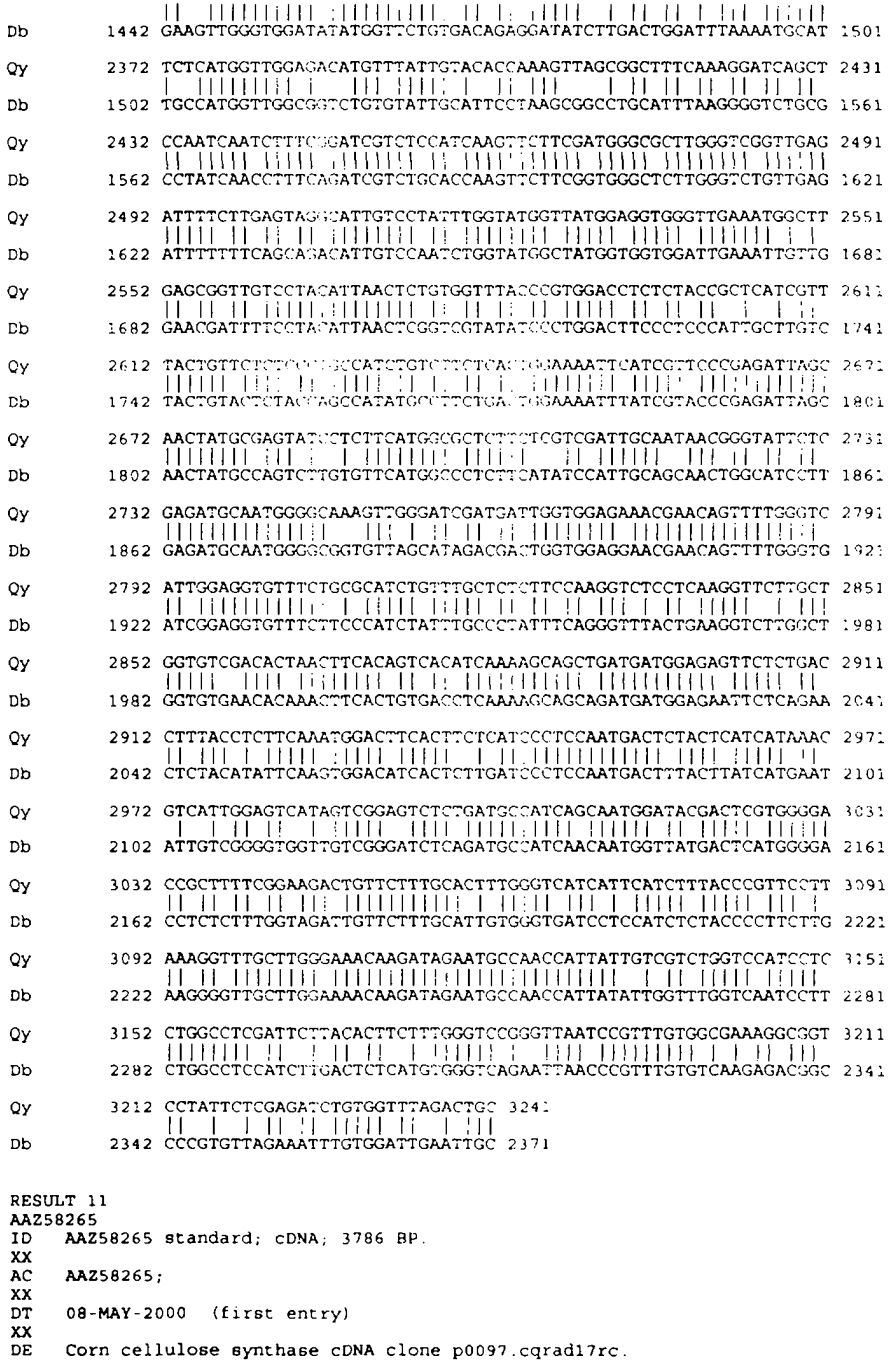
Qy       1115 ACTGTCTTGTCTATTCTTGCTGTTGATTATCCTGTCGATAAGGTTGCTTGTTACGTATCT 1174

Db        248 ACTGTTCTATCTATCCTTGCTGTTGATTATCCAGTTGATAAAGTTGCATGCTATGTCTCA 307

Qy       1175 GATGATGGTGCTGCTATGCTTACTTTCGAAGCTCTTTCTGAGACCGCTGAATTCGCAAGG 1234

Db        308 GATGATGGTGCTGCTATGCTTACTTTTGAAGCACTGTCTGAGACATCTGAATTTGCTAGG 367
```





RESULT 11
AAZ58265
ID AAZ58265 standard; cDNA; 3786 BP.
XX
AC AAZ58265;
XX
DT 08-MAY-2000 (first entry)
XX
DE Corn cellulose synthase cDNA clone p0097.cqrad17rc.

```
XX
KW   Corn; maize; cellulose synthase; transgenic plant; ss.
XX
OS   Zea mays.
XX
FH   Key             Location/Qualifiers
FT   CDS             2..3499
FT                   /*tag= a
FT                   /partial
XX
PN   WO200004166-A2.
XX
PD   27-JAN-2000.
XX
PF   13 JUL-1999;   99WO-US15871.
XX
PR   14-JUL-1998;   98US-0092844.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Allen SM,  Fader GM,  Falco SC,  Kinney AJ,  Lightner JE,  Miao G;
PI   Rafalski JA,  Thorpe CJ;
XX
DR   WPI; 2000-182431/16.
DR   P-PSDB; AAY58834.
XX
PT   Nucleic acid fragments encoding cellulose biosynthetic enzyme useful as
PT   probes for isolating cDNAs and genes encoding homologous proteins, for
PT   producing transgenic plants
XX
PS   Claim 9; Page 44-45; 93pp; English.
XX
CC   The present sequence is that of cDNA clone p0097.cgrad17rc encoding
CC   a portion (see AAY58833) of corn cellulose synthase (CS).  The cDNA
CC   clone was isolated from a european corn borer 4 times-infected corn
CC   stage V9 whorl section cDNA library on the basis of homology to
CC   Arabidopsis and cotton CS sequences.  The invention relates to
CC   isolated nucleic acid fragments encoding plant CS and to CS
CC   polypeptides.  It also relates to the construction of a chimeric
CC   gene encoding all or a portion of the CS, in sense or antisense
CC   orientation, where expression of the gene results in altered levels
CC   of the CS in transformed host cells.  The host cells can be used to
CC   screen compounds for their ability to inhibit CS activity.  CS
CC   nucleic acids are also useful for producing transgenic plants having
CC   altered levels of CS, and hence altered levels of fibre.  CS may
CC   also serve as a target for the development of novel herbicides.
XX
SQ   Sequence 3786 BP; 846 A; 967 C; 1111 G; 862 T; 0 other;

  Query Match           41.1%;  Score 1416.4;  DB 21;  Length 3786;
  Best Local Similarity  66.7%;  Pred. No. 0;
  Matches 2128;  Conservative    0;  Mismatches 1026;  Indels   36;  Gaps    6;

Qy        103 GACATGTCAAATCTGCAGAGATGAGATCGAATTGACTGTTGATGGAGAACCGTTTGTGGC 162
              | | || || || ||| | || ||| |||   ||    | || || || || || |||||
Db        331 GCCGTGCCAGATATGCGGCGACGAGGTCGGGGTGGGCTTCGACGGGGAGCCCTTCGTGGC 390

Qy        163 ATGTAACGAATGTGCATTCCCTGTGTGTAGACCTTGCTATGAGTACGAAAGACGAGAAGG 222
                {{ ||||| || || ||||| |, ||  |  | |||| ||||||||  | || || ||
Db        391 GTGCAACGAGTGCGCCTTCCCCGTCTGCCGCGCCTGCTACGAGTACGAGCGCCGCGAGGG 450

Qy        223 CAATCAAGCTTGTCCACAGTGCAAAACCCGTTTCAAACGTCTTAAAGGAAGTCCAAGAGT 282
              |   |||| || || |||||||  |||| | ||| || || || ||  | ||  | ||
Db        451 CTCGCAAGCGTGCCCGCAGTGCAGGACCCGCTACAAGCGCCTCAAGGGCTGCCCGCGGGT 510

Qy        283 TGAAGGTGATGAAGAGGAAGATGACATTGATGATTTAGACAATGAGTT---------TGA 333
              |  || || || ||||| || | | || ||  | ||      ||||           ||
Db        511 GGCCGGCGACGAGGAGGAGGACGGCGTCGACGACCTGGAGGGCGAGTTCGGCCTGCAGGA 570

Qy        334 GTATGGAAATAATGGGATTGGATTTGATCAGGTTTCTGAAGGTATGTCAATCTCTCGTCG 393
                  |      | | |   |      |  | ||  | |.    ||   |   | |
Db        571 CGGCGCCGCCCACGAGGACGACCCGCAGTACGTCGCCGAGTCCATGCTCAGGGCGCAGAT 630

Qy        394 CAACTCCGGTTTCCCACAATCTGATTTGGATTCAGCTCCACCTGGCTCTCAGATTCCATT 453
              | || |||   |           |                    || | | |  |  | || |
Db        631 GAGCTACGGCCGCGGCGGCGACGCGCACCCCGGCTTCAGCCCCGTCCCCAACGTGCCGCT 690

Qy        454 GCTGACTTACGGCGACGAGGACGTTGAGATTTCTTCTGATAGACATGCTCTTATTGTTCC 513
              || ||  ||||| |   ||  | ||| ||  |  | ||   || || ||  |     |
Db        691 CCTCACCAACGGCCAGATGGTTGATGACATCCCGCCGGAGCAGCACGCGCTCGTGCCGTC 750

Qy        514 TCCTTCACTTGGTGGTCAT------GGCAATAGAGTTCATCCTGTTTCTCTTTCTGACCC 567
                    || ||          ||||| ||  | || ||  |  || |  | || ||
Db        751 CTACATGAGCGGCGGCGGCGGCGGGGGCAAGAGGATCCACCCGCTCCCTTTCGCAGATCC 810

Qy        568 GACCGTGGCTGCACATCGAAGGCTGATGGTACCTCAGAAAGATCTTGCGGTTTATGGTTA 627
              | | |  | |  || |  ||     |||| ||     || ||||| || |  || || ||
Db        811 CAACCTTCCAGTGCAACCGAGATCCATGGACCCGTCCAAGGATCTGGCCGCCTACGGATA 870

Qy        628 TGGAAGTGTCGCTTGGAAAGATCGGATGGAGGAATGGAAGAGAAAGCAGAATGAGAAACT 687
              ||| || || || ||||| ||  | |||||||  ||||||   ||||||   |||   ||
Db        871 TGGCAGCGTGGCCTGGAAGGAGAGAATGGAGGGCTGGAAGCAGAAGCAG---GAGCGCCT 927

Qy        688 TCAGGTTGTTAGGCATGAAGGAGATCCTGATTTTGAAGATGGTGATGATGCTGATTTTCC 747
              |||  ||| |||   || || |    |||||     |||||  || |||||  ||| | ||
Db        928 GCAGCATGTCAGGAGCGAGGGTGGCGGTGATTGG---GATGGCGACGATGCAGATCTGCC 984

Qy        748 AATGATGGATGAGGGAAGGCAGCCATTGTCTATGAAGATACCAATCAAATCGAGCAAGAT 807
              | | |||||||| |  |||||||||||||||| |  ||  | ||  ||   |||  ||
Db        985 ACTAATGGATGAAGCTAGGCAGCCATTGTCCAGAAAAGTCCCTATATCATCAAGCCGAAT 1044

Qy        808 AAATCCTTACCGGATGTTAATTGTGCTACGTCTTGTGATTCTTGGTCTCTTCTTTCACTA 867
              |||||  ||| |||||  | || ||  | ||  | ||| || |  |||  |||||  |||||
Db       1045 TAATCCCTACAGGATGATTATCGTTATCCGGTTGGTGGTTTTGGGTTTCTTCTTCCACTA 1104

Qy        868 CCGTATTCTTCACCCCGTCAAAGATGCATATGCTTTGTGGCTTATTTCTGTTATATGTGA 927
              |||  |  | || || |  ||||||||||  ||| ||||||||  || |||||  || |||||
Db       1105 CCGAGTGATGCATCCGGCGAAAGATGCATTTGCATTGTGGCTCATATCTGTAATCTGTGA 1164

Qy        928 GATATGGTTTGCTGTTTCATGGGTTCTTGATCAGTTCCCTAAATGGTACCCTATCGAGCG 987
              || ||||||||  | || ||| ||||||||||||||||| || |||   || ||||||  |
Db       1165 AATCTGGTTTGCGATGTCCTGGATTCTTGATCAGTTCCCAAAGTGGCTTCCAATCGAGAG 1224

Qy        988 AGAAACGTACTTGGACCGACTCTCATTAAGATATGAGAAAGAAGGGAAACCGTCGGGACT 1047
              ||| || ||| ||||||  | ||| |||| | || || |||||  |||| ||   ||
Db       1225 AGAGACTTACCTGGACCGTTTGTCACTAAGGTTTGACAAGGAAGGTCAACCCTCTCAGCT 1284

Qy       1048 ATCCCCTGTGGATGTATTTGTTAGTACAGTGGATCCATTGAAAGAGCCTCCGCTTATTAC 1107
              | ||  | ||  | ||||| ||||| || |||||     || || |||||  |  | ||
Db       1285 TGCTCCAATCGACTTCTTTGTCAGTACGGTTGATCCCACAAAGGAACCTCCCTTGGTCAC 1344

Qy       1108 TGCAAATACTGTCTTGTCTATTCTTGCTGTTGATTATCCTGTCGATAAGGTTGCTTGTTA 1167
              || || |||||  | || || ||| |||| |||||||| || || |||||  | || ||
Db       1345 AGCGAACACTGTCCTTTCCATCCTTTCTGTGGATTATCCGGTTGAGAAGGTCTCCTGCTA 1404

Qy       1168 CGTATCTGATGATGGTGCTGCTATGCTTACTTTCGAAGCTCTTTCTGAGACCGCTGAATT 1227
              || |||||||||||||||||| |||||||| || |||||  | |||||  ||  |||||||
Db       1405 TGTTTCTGATGATGGTGCTGCAATGCTTACGTTTGAAGCATTGTCTGAAACATCTGAATT 1464

Qy       1228 CGCAAGGAAATGGGTTCCTTTCTGCAAGAAATATTGTATTGAGCCTCGTGCTCCCGAATG 1287
              |||| ||||||||||||||||| |||| || | |  ||| ||||||||||||||| || ||
Db       1465 TGCAAAGAAATGGGTTCCTTTCAGCAAAAAGTTTAATATCGAGCCTCGTGCTCCTGAGTG 1524

Qy       1288 GTATTTCTGCCATAAAATGGACTACTTGAAGAATAAAGTTCATCCCGCATTTGTTAGGGA 1347
              ||| |||   || || || |||||| ||||  | || |||  | |  ||||||||||||||
Db       1525 GTACTTCCAACAGAAGATAGACTACCTGAAAGACAAGGTTGCTGCTTCATTTGTTAGGGA 1584

Qy       1348 GCGGCGAGCCATGAAGAGAGATTATGAAGAATTCAAAGTAAAGATCAATGCTTTAGTAGC 1407
              | || | || |||||||||||| || |||||||| |||| |||||||| || || ||
Db       1585 GAGGAGGGCGATGAAGAGAGAATACGAGGAATTCAAGGTAAGGATCAATGCCTTGGTTGC 1644

Qy       1408 AACAGCACAGAAAGTGCCTGAGGATGGTTGGACTATGCAAGACGGTACACCTTGGCCCGG 1467
              || ||| || || || |||||||| || ||||| ||||||| || |  || ||||| ||
Db       1645 AAAAGCCCAAAAGGTTCCTGAGGAAGGATGGACAATGCAAGATGGAAGCCCCTGGCCTGG 1704

Qy       1468 TAATAGTGTGCGAGATCATCCTGGCATGATTCAGGTCTTCCTTGGAAGTGACGGTGTTCG 1527
```

```
b       1705 AAACAACGTACGCGATCATCCTGGAATGATTCAGGTATTCCTTGGCCAAAGTGGCGGTCG 1764
y       1528 TGATGTCGAAAACAACGAGTTGCCTCGATTAGTTTACGTTTCTCGTGAGAAGAGACCCGG 1587
b       1765 TGATGTGGAAGGAAATGAGTTGCCTCGCCTGGTTTATGTCTCGAGAGAAAAGAGGCCAGG 1824
y       1588 ATTTGATCACCATAAGAAGGCTGGAGCTATGAATTCCCTGATACGAGTCTCTGGGGTTCT 1647
b       1825 TTATAACCATCACAAGAAGGCTGGTGCCATGAATGCACTGGTCCGTGTCTCTGCTGTCTT 1884
y       1648 ATCAAATGCTCCTTACCTTCTGAATGTCGATTGTGATCACTACATCAACAATAGCAAAGC 1707
b       1885 ATCAAATGCTGCATACCTATTGAACTTGGACTGTGATCACTACATCAACAATAGCAAGGC 1944
y       1708 TCTTAGAGAAGCAATGTGTTTCATGATGGATCCTCAGTCAGGAAAGAAAATCTGTTATGT 1767
b       1945 CATAAAAGAGGCTATGTGTTTCATGATGGATCCTTTGGTGGGGAAGAAAGTGTGCTATGT 2004
y       1768 TCAGTTCCCTCAAAGGTTCGATGGGATTGATAGGCACGATCGATACTCAAATCGCAATGT 1827
b       2005 ACAGTTCCCTCAGAGGTTTGATGGTATTGACAAAAATGATCGATACGCTAACAGGAACGT 2064
y       1828 TGTGTTCTTTGATATCAATATGAAAGGTTTGGATGGGCTACAAGGGCCTATATACGTCGG 1887
b       2065 TGTCTTTTTTGACATCAACATGAAAGGTTTGGACGGTATTCAAGGACCCATTTATGTGGG 2124
y       1888 TACAGGTTGTGTTTTCAGGAGGCAAGCGCTTTACGGATTTGATGCACCGAAGAAGAAGAA 1947
b       2125 TACTGGATGTGTTTTCAGACGGCAGGCACTGTATGGTTATGATGCTCCTAAAACGAAGAA 2184
y       1948 GGGCCCACGTAAGACATGCAATTGCTGGCCAAAATGGTGTCTCCTATGTTTTGGTTCAAG 2007
b       2185 GCCACCATCAAGAACTTGCAACTGCTGGCCCAAGTGGTGCCTCTCTTGCTGCTGCAGCAG 2244
y       2008 AAAGAATCGTAAAGCAAAGACAGTGGCTGCGGATAAGAAGAAGAAGAATAG--------- 2058
b       2245 GAACAAGAATAAAAAGAAGACTACAAAACCAAAGACGGAGAAGAAGAAAAGATTATTTTT 2304
y       2059 ---GGAAGCGTCAAAGCAGATCCACGCATTAGAAAAATATCGAAGAGGGCCGCGGTCATAA 2115
b       2305 CAAGAAAGCAGAAAACCCATCTCCTGCATATGCTTTGGGTGAAATTGATGAAGGTGCTCC 2364
y       2116 AGTTCTTAACGTAGAACAGTCAACCGAGGCAATGCAAATGAAGTTGCAGAAGAAATATGG 2175
b       2365 AGGTGCTGATATCGAGAAGGCCGGAATCGTAAATCAACAGAAACTAGAGAAGAAATTTGG 2424
y       2176 GCAGTCTCCTGTATTTGTTGCATCTGCGCGTCTGGAGAATGGTGGGATGGCTAGAAACGC 2235
b       2425 GCAGTCTTCTGTTTTTGTCGCATCAACACTTCTTGAGAACGGAGGGACCCTGAAGAGCGC 2484
y       2236 AAGCCCGGCTTGTCTGCTTAAAGAAGCCATCCAAGTCATTAGTCGCGGATATGAAGATAA 2295
b       2485 AAGTCCAGCTTCTCTTCTGAAGGAAGCTATACATGTTATCAGCTGCGGCTACGAAGACAA 2544
y       2296 AACTGAATGGGGAAAAGAGATTGGGTGGATCTATGGTTCTGTTACCGAAGATATTCTTAC 2355
b       2545 GACCGACTGGGGAAAAGAGATTGGCTGGATTTACGGATCGATCACAGAGGATATCTTGAC 2604
y       2356 GGGTTCTAAGATGCATTCTCATGGTTGGAGACATGTTTATTGTACACCAAAGTTAGCGGC 2415
b       2605 TGGATTTAAGATGCACTGCCATGGCTGGCGGTCTATTTACTGCATCCCGAAGCGGCCTGC 2664
y       2416 TTTCAAAGGATCAGCTCCAATCAATCTTTCGGATCGTCTCCATCAAGTTCTTCGATGGGC 2475
b       2665 ATTCAAAGGTTCTGCGCCTCTGAACCTTTCCGACCGTCTTCACCAGGTCCTTCGCTGGGC 2724
y       2476 GCTTGGGTCGGTTGAGATTTTCTTGAGTAGGCATTGTCCTATTTGGTATGGTTATGGAGG 2535
b       2725 CCTTGGGTCCGTCGAAATTTTCTTCAGCAAGCACTGCCCACTTTGGTACGGATACGGCGG 2784
y       2536 TGGGTTGAAATGGCTTGAGCGGTTGTCCTACATTAACTCTGTGGTTTACCCGTGGACCTC 2595
```

```
Db      2785 CGGGCTAAAATTC[illegible]GGAAAGGTTTTCTTATATCAACTCCATCGTTTATCCCTGGACGTC 2844
Qy      2596 TCTACCGCTCATC[illegible]TACTGTTCTCTCCCTGCCATCTGTCTTCTCACTGGAAAATTCAT 2655
Db      2845 CATTCCTCTCCTG[illegible]TTACTGTACCTTGCCTGCCATCTGCCTGCTCACGGGGAAGTTTAT 2904
Qy      2656 CGTTCCCGAGATTA[illegible]AACTATGCGAGTATCCTCTTCATGGCGCTCTTCTCGTCGATTGC 2715
Db      2905 CACACCAGAGCTTA[illegible]AATGTCGCCAGTATCTGGTTCATGGCACTTTTCATCTGCATCTC 2964
Qy      2716 AATAACGGGTATT[illegible]GAGATGCAATGGGGCAAAGTTGGGATCGATGATTGGTGGAGAAA 2775
Db      2965 CGTGACCGGCAT[illegible]GGAAATGAGGTGGAGTGGCGTGGCCATCGACGACTGGTGGAGGAA 3024
Qy      2776 CGAACAGTTTTG[illegible]ATTGGAGGTGTTTCTGCGCATCTGTTTGCTCTCTTCCAAGGTCT 2835
Db      3025 CGAGCAGTT[illegible]ATGGAG[illegible]ATCTGTTCGCGGTGTTCCAGGGCCT 3084
Qy      2836 CCTCAAGGTTCTT[illegible]AACAGTCACATCAAAAGCAGCTGATGA 2895
Db      3085 GCTGAAGGTGTTC[illegible]GCATCGACACGAGCTTCACCGTGACGTCGAAGGCCGGGGACGA 3144
Qy      2896 TGGAGAGTTCTCTGACCTTTACCTCTTCAAATGACTTCACTTCTCATCCCTCCAATGAC 2955
Db      3145 CGAGGAGTTCTCGGAGCTGTACACGTTCAAGT[illegible]GACCACCCTGCTGATACCCCCGACCAC 3204
Qy      2956 TCTACTCATCATAAACGTCATTGGAGTCATAGTCGGAGTCTCTGATGCCATCAGCAATGG 3015
Db      3205 GCTCCTCCTGCTGAACTTCATCGGGGTGGTGG[illegible]CGGGATCTCGAACGCGATCAACAACGG 3264
Qy      3016 ATACGACTCGTGGG[illegible]ACCGCTTTTCGGAAGAC[illegible]GTTCTTTGCACTTTGGGTCATCATTCA 3075
Db      3265 GTACGAGTCGTGGGGCCCCCTGTTCGGGAAGC[illegible]CTTCTTCGCCTTCTGGGTGATCGTCCA 3324
Qy      3076 TCTTTACCCGTTCCTTAAAGGTTTGCTTGGGAAACAAGATAGAATGCCAACCATTATTGT 3135
Db      3325 CCTGTACCCGTTCCTCAAGGGTCTGGTGGGGAGGCAGAACAGGACGCCGACGATCGTCAT 3384
Qy      3136 CGTCTGGTCCATCCTCCTGGCCTCGATTCTTA[illegible]ACTTCTTTGGGTCCGGGTTAATCCGTT 3195
Db      3385 CGTCTGGTCCATCCTGCTGGCCTCGATCTTCT[illegible]GCTCCTGTGGGTCCGCGTCGACCCGTT 3444
Qy      3196 TGTGGCGAAAGG-- CGGTCCTATTCTCGAGA[illegible]CTGTGGTTTAGACTGCTTGTGATTCGA 3252
Db      3445 CCTCGCCAAGAGCAACGGCCCGCTCCTGGAGGAGTGTGGCCTGGACTGCAACTGAAGTGG 3504
Qy      3253 TTGACCGGTG 3262
Db      3505 GGGCCCCCTG 3514
```

```
RESULT 12
AAZ99509
ID   AAZ99509 standard; DNA; 3813 BP.
XX
AC   AAZ99509;
XX
DT   03-JUL-2000  (first entry)
XX
DE   DNA encoding a maize cellulose synthase.
XX
KW   Maize; cellulose synthase; stalk quality; stand; silage; cellulose;
KW   transgenic plant; plant breeding marker; ss.
XX
OS   Zea mays.
XX
FH   Key             Location/Qualifiers
FT   CDS             215..3493
FT                   /*tag= a
FT                   /product= "cellulose synthase"
FT                   /note= "no termination codon given"
XX
PN   WO200009706-A2.
```

XX
PD 24-FEB-2000.
XX
PF 16-AUG-1999; 99WO-US18760.
XX
PR 17-AUG-1998; 98US-0096822.
XX
PA (PION-) PIONEER HI-BRED INT INC.
XX
PI Dhugga KS, Helentjaris TG, Bowen BA, Wang X;
XX
DR WPI; 2000-224343/19.
DR P-PSDB; AAY84113.
XX
PT New genes which encode maize cellulose synthase polypeptides in plants
PT useful for modulating the expression of cellulose synthase in plants
PT and to produce transgenic plants expressing the novel protein -
XX
PS Claim 1; Page 129-134; 119pp; English.
XX
CC The present sequence encodes a maize cellulose synthase polypeptide.
CC The cellulose synthase can be used for the improvement of stalk quality
CC for improved stand or silage. It also provides an increased concentration
CC of cellulose in the pericarp, hardening the kernel and improving its
CC handling ability. The sequences are used to produce transgenic plants
CC and seeds expressing the cellulose synthase. The polynucleotide is
CC used for modulating, preferably increasing, the level of the synthase
CC in a plant cell. The plants are preferably monocots. The polynucleotide
CC is also used as a probe or primer in the detection quantitation or
CC isolation of gene transcripts. The probes are useful in detecting
CC deficiencies in the level of mRNA in screenings for desired transgenic
CC plant, for detecting mutations in the gene, for monitoring upregulation
CC of expression or changes in enzyme activity in screening assays of
CC compounds, for detection of any number of allelic variants of the gene,
CC or for use as molecular markers in plant breeding programs. The
CC isolated nucleic acids of the present invention can also be used for
CC recombinant expression of their encoded polypeptides or for use as
CC immunogens in the preparation and/or screening of antibodies. The
CC proteins can be employed in assays for enzyme agonists or antagonists
CC of enzyme function or for use of immunogens or antigens to obtain
CC antibodies specifically immunoreactive with a protein.
XX
SQ Sequence 3813 BP; 873 A; 967 C; 1111 G; 862 T; 0 other;

Query Match 41.1%; Score 1416.4; DB 21; Length 3813;
Best Local Similarity 66.7%; Pred. No. 0;
Matches 2128; Conservative 0; Mismatches 1026; Indels 36; Gaps 6;

```
Qy        103 GACATGTCAAATCTGCAGAGATGAGATCGAATTGACTGTTGATGGAGAACCGTTTGTGGC 162
Db        331 GCCGTGCCAGATATGCGGCGACGAAGTCGGGTGGGCTTC   GGGAGCCCTTCGTGGC 390

Qy        163 ATGTAACGAATGTGCATTCCCTGTGTGTAGACCTTGCTAT   TACGAAAGACGAGAAGG 222
Db        391 GTGCAACGAGTGCGCCTTCCCCGTCTGCCGCGCCTGCTACGAGTACGAGCGCCGCGAGGG 450

Qy        223 CAATCAAGCTTGTCCACAGTGCAAAACCCGTTTCAAACGTC   AAAGGAAGTCCAAGAGT 282
Db        451 CTCGCAAGCGTGCCCGCAGTGCAGGACCCGCTACAAGCGCC   AAGGGCTGCCCGCGGGT 510

Qy        283 TGAAGGTGATGAAGAGGAAGATGACATTGATGATTTAGACAA GAGTT---------TGA 333
Db        511 GGCCGGCGACGAGGAGGAGGACGGCGTCGACGACCTGGAGG GAGTTCGGCCTGCAGGA 570

Qy        334 GTATGGAAATAATGGGATTGGATTTGATCAGGTTTCTGAAG TATGTCAATCTCTCGTCG 393
Db        571 CGGCGCCGCCCACGAGGACGACCCGCAGTACGTCGCCGAGT CATGCTCAGGGCGCAGAT 630

Qy        394 CAACTCCGGTTTCCCACAATCTGATTTGGATTCAGCTCCAC GGCTCTCAGATTCCATT 453
Db        631 GAGCTACGGCCGCGGCGGCGACGCGCACCCCGGCTTCAGCC GTCCCCAACGTGCCGCT 690

Qy        454 GCTGACTTACGGCGACGAGGACGTTGAGATTTCTTCTGATA ACATGCTCTTATTGTTCC 513

Db        691 CCTCACCAACGGCCAGATGGTTGATGACATCCCGCCGGAGCAGCACGCGCTCGTGCCGTC 750

Qy        514 TCCTTCACTTGGTGGTCAT------GGCAATAGAGTTCATCCTGTTTCTCTTTCTGACCC 567
Db        751 CTACATGAGCGGCGGCGGCGGGGGGCAAGAGGATCCACCCGCTCCCTTTCGCAGATCC 810

Qy        568 GACCGTGGCTGCACATCGAAGGCTGATGGTACCTCAGAAAGATCTTGCGGTTTATGGTTA 627
Db        811 CAACCTTCCAGTGCAACCGAGATCCATGGACCCGTCCAAGGATCTGGCCGCCTACGGATA 870

Qy        628 TGGAAGTGTCGCTTGGAAAGATCGGATGGAGGAATGGAAGAGAAAGCAGAATGAGAAACT 687
Db        871 TGGCAGCGTGGCCTGGAAGGAGAGAATGGAGGGCTGGAAGCAGAAGCAG---GAGCGCCT 927

Qy        688 TCAGGTTGTTAGGCATGAAGGAGATCCTGATTTTGAAGATGGTGATGATGCTGATTTTCC 747
Db        928 GCAGCATGTCAGGAGCGAGGGTGGCGGTGATTGG---GATGGCGACGATGCAGATCTGCC 984

Qy        748 AATGATGGATGAGGGAAGGCAGCCATTGTCTATGAAGATACCAATCAAATCGAGCAAGAT 807
Db        985 ACTAATGGATGAAGCTAGGCAGCCATTGTCCAGAAAAGTCCCTATATCATCAAGCCGAAT 1044

Qy        808 AAATCCTTACCGGATGTTAATTGTGCTACGTCTTGTGATTCTTGGTCTCTTCTTTCACTA 867
Db        1045 TAATCCCTACAGGATGATTATCGTTATCCGGTTGGTGGTTTTGGGTTTCTTCTTCCACTA 1104

Qy        868 CCGTATTCTTCACCCCGTCAAAGATGCATATGCTTTGTGGCTTATTTCTGTTATATGTGA 927
Db        1105 CCGAGTGATGCATCCGGCGAAAGATGCATTTGCATTGTGGCTCATATCTGTAATCTGTGA 1164

Qy        928 GATATGGTTTGCTGTTTCATGGGTTCTTGATCAGTTCCCTAAATGGTACCCTATCGAGCG 987
Db        1165 AATCTGGTTTGCGATGTCCTGGATTCTTGATCAGTTCCCAAAGTGGCTTCCAATCGAGAG 1224

Qy        988 AGAAACGTACTTGGACCGACTCTCATTAAGATATGAGAAAGAAGGGAAACCGTCGGGACT 1047
Db        1225 AGAGACTTACCTGGACCGTTTGTCACTAAGGTTTGACAAGGAAGGTCAACCCTCTCAGCT 1284

Qy        1048 ATCCCCTGTGGATGTATTTGTTAGTACAGTGGATCCATTGAAAGAGCCTCCGCTTATTAC 1107
Db        1285 TGCTCCAATCGACTTCTTTGTCAGTACGGTTGATCCCACAAAGGAACCTCCCTTGGTCAC 1344

Qy        1108 TGCAAATACTGTCTTGTCTATTCTTGCTGTTGATTATCCTGTCGATAAGGTTGCTTGTTA 1167
Db        1345 AGCGAACACTGTCCTTTCCATCCTTTCTGTGGATTATCCGGTTGAGAAGGTCTCCTGCTA 1404

Qy        1168 CGTATCTGATGATGGTGCTGCTATGCTTACTTTCGAAGCTCTTTCTGAGACCGCTGAATT 1227
Db        1405 TGTTTCTGATGATGGTGCTGCAATGCTTACGTTTGAAGCATTGTCTGAAACATCTGAATT 1464

Qy        1228 CGCAAGGAAATGGGTTCCTTTCTGCAAGAAATATTGTATTGAGCCTCGTGCTCCCGAATG 1287
Db        1465 TGCAAAGAAATGGGTTCCTTTCAGCAAAAAGTTTAATATCGAGCCTCGTGCTCCTGAGTG 1524

Qy        1288 GTATTTCTGCCATAAAATGGACTACTTGAAGAATAAAGTTCATCCCGCATTTGTTAGGGA 1347
Db        1525 GTACTTCCAACAGAAGATAGACTACCTGAAAGACAAGGTTGCTGCTTCATTTGTTAGGGA 1584

Qy        1348 GCGGCGAGCCATGAAGAGAGATTATGAAGAATTCAAAGTAAAGATCAATGCTTTAGTAGC 1407
Db        1585 GAGGAGGGCGATGAAGAGAGAATACGAGGAATTCAAGGTAAGGATCAATGCCTTGGTTGC 1644

Qy        1408 AACAGCACAGAAAGTGCCTGAGGATGGTTGGACTATGCAAGACGGTACACCTTGGCCCGG 1467
Db        1645 AAAAGCCCAAAAGGTTCCTGAGGAAGGATGGACAATGCAAGATGGAAGCCCCTGGCCTGG 1704

Qy        1468 TAATAGTGTGCGAGATCATCCTGGCATGATTCAGGTCTTCCTTGGAAGTGACGGTGTTCG 1527
Db        1705 AAACAACGTACGCGATCATCCTGGAATGATTCAGGTATTCCTTGGCCAAAGTGGCGGTCG 1764

Qy        1528 TGATGTCGAAAACAACGAGTTGCCTCGATTAGTTTACGTTTCTCGTGAGAAGAGACCCGG 1587
Db        1765 AAACAACGTACGCGATCATCCTGGAATGATTCAGGTATTCCTTGGCCAAAGTGGCGGTCG 1824
```

```
b      1765 TGATGTGGAAGGAAATGAGTTGCCTCGCCTGGTTTATGTCTCGAGAGAAAAGAGGCCAGG 1824

y      1588 ATTTGATCACCATAAGAAGGCTGGAGCTATGAATTCCCTGATACGAGTCTCTGGGGTTCT 1647

b      1825 TTATAACCATCACAAGAAGGCTGGTGCCATGAATGCACTGGTCCGTGTCTCTGCTGTCTT 1884

y      1648 ATCAAATGCTCCTTACCTTCTGAATGTCGATTGTGATCACTACATCAACAATAGCAAAGC 1707

b      1885 ATCAAATGCTGCATACCTATTGAACTTGGACTGTGATCACTACATCAACAATAGCAAGGC 1944

y      1708 TCTTAGAGAAGCAATGTGTTTCATGATGGATCCTCAGTCAGGAAAGAAAATCTGTTATGT 1767

b      1945 CATAAAAGAGGCTATGTGTTTCATGATGGATCCTTTGGTGGGGAAGAAAGTGTGCTATGT 2004

y      1768 TCAGTTCCCTCAAAGGTTCGATGGGATTGATAGGCACGATCGATACTCAAATCGCAATGT 1827

b      2005 ACAGTTCCCTCAGAGGTTTGATGGTATTGACAAAAATGATCGATACGCTAACAGGAACGT 2064

y      1828 TGTGTTCTTTGATATCAATATGAAAGGTTTGGATGGGCTACAAGGGCCTATATACGTCGG 1887

b      2065 TGTCTTTTTTGACATCAACATGAAAGGTTTGGACGGTATTCAAGGACCCATTTATGTGGG 2124

y      1888 TACAGGTTGTGTTTTCAGGAGGCAAGCGCTTTACGGATTTGATGCACCGAAGAAGAAGAA 1947

b      2125 TACTGGATGTGTTTTCAGACGGCAGGCACTGTATGGTTATGATGCTCCTAAAACGAAGAA 2184

y      1948 GGGCCCACGTAAGACATGCAATTGCTGGCCAAAATGGTGTCTCCTATGTTTTGGTTCAAG 2007

b      2185 GCCACCATCAAGAACTTGCAACTGCTGGCCCAAGTGGTGCCTCTCTTGCTGCTGCAGCAG 2244

y      2008 AAAGAATCGTAAAGCAAAGACAGTGGCTGCGGATAAGAAGAAGAAGAATAG--------- 2058

b      2245 GAACAAGAATAAAAAGAAGACTACAAAACCAAAGACGGAGAAGAAGAAAAGATTATTTTT 2304

y      2059 ---GGAAGCGTCAAAGCAGATCCACGCATTAGAAAATATCGAAGAGGGCCGCGGTCATAA 2115

b      2305 CAAGAAAGCAGAAAACCCATCTCCTGCATATGCTTTGGGTGAAATTGATGAAGGTGCTCC 2364

y      2116 AGTTCTTAACGTAGAACAGTCAACCGAGGCAATGCAAATGAAGTTGCAGAAGAAATATGG 2175

b      2365 AGGTGCTGATATCGAGAAGGCCGGAATCGTAAATCAACAGAAACTAGAGAAGAAATTTGG 2424

y      2176 GCAGTCTCCTGTATTTGTTGCATCTGCGCGTCTGGAGAATGGTGGGATGGCTAGAAACGC 2235

b      2425 GCAGTCTTCTGTTTTTGTCGCATCAACACTTCTTGAGAACGGAGGGACCCTGAAGAGCGC 2484

y      2236 AAGCCCGGCTTGTCTGCTTAAAGAAGCCATCCAAGTCATTAGTCGCGGATATGAAGATAA 2295

b      2485 AAGTCCAGCTTCTCTTCTGAAGGAAGCTATACATGTTATCAGCTGCGGCTACGAAGACAA 2544

y      2296 AACTGAATGGGGAAAAGAGATTGGGTGGATCTATGGTTCTGTTACCGAAGATATTCTTAC 2355

b      2545 GACCGACTGGGGAAAAGAGATTGGCTGGATTTACGGATCGATCACAGAGGATATCTTGAC 2604

y      2356 GGGTTCTAAGATGCATTCTCATGGTTGGAGACATGTTTATTGTACACCAAAGTTAGCGGC 2415

b      2605 TGGATTTAAGATGCACTGCCATGGCTGGCGGTCTATTTACTGCATCCCGAAGCGGCCTGC 2664

y      2416 TTTCAAAGGATCAGCTCCAATCAATCTTTCGGATCGTCTCCATCAAGTTCTTCGATGGGC 2475

b      2665 ATTCAAAGGTTCTGCGCCTCTGAACCTTTCCGACCGTCTTCACCAGGTCCTTCGCTGGGC 2724

y      2476 GCTTGGGTCGGTTGAGATTTTCTTGAGTAGGCATTGTCCTATTTGGTATGGTTATGGAGG 2535

b      2725 CCTTGGGTCCGTCGAAATTTTCTTCAGCAAGCACTGCCCACTTTGGTACGGATACGGCGG 2784

y      2536 TGGGTTGAAATGGCTTGAGCGGTTGTCCTACATTAACTCTGTGGTTTACCCGTGGACCTC 2595

b      2785 CGGGCTAAAATTCCTGGAAAGGTTTTCTTATATCAACTCCATCGTTTATCCCTGGACGTC 2844

y      2596 TCTACCGCTCATCGTTTACTGTTCTCTCCCTGCCATCTGTCTTCTCACTGGAAAATTCAT 2655

b      2845 CATTCCTCTCCTGGCTTACTGTACCTTGCCTGCCATCTGCCTGCTCACGGGGAAGTTTAT 2904
```

```
Qy     2656 CGTTCCCGAGATTAGCAACTATGCGAGTATCCTCTTCATGGCGCTCTTCTCGTCGATTGC 2715

Db     2905 CACACCAGAGCTTACCAATGTCGCCAGTATCTGGTTCATGGCACTTTTCATCTGCATCTC 2964

Qy     2716 AATAACGGGTATTCTCGAGATGCAATGGGGCAAAGTTGGGATCGATGATTGGTGGAGAAA 2775

Db     2965 CGTGACCGGCATCCTGGAAATGAGGTGGAGTGGCGTGGCCATCGACGACTGGTGGAGGAA 3024

Qy     2776 CGAACAGTTTTGGGTCATTGGAGGTGTTTCTGCGCATCTGTTTGCTCTCTTCCAAGGTCT 2835

Db     3025 CGAGCAGTTCTGGGTCATCGGAGGCGTTTCGGCGCATCTGTTCGCGGTGTTCCAGGGCCT 3084

Qy     2836 CCTCAAGGTTCTTGCTGGTGTCGACACTAACTTCACAGTCACATCAAAAGCAGCTGATGA 2895

Db     3085 GCTGAAGGTGTTCTGGGCATCGACACGAGCTTCACCGTGACGTCGAAGGCCGGGGACGA 3144

Qy     2896 TGGAGAGTTCTCTGAGCTTTACTTCTTCAAATGGACTTCACTTCTCATCCCTCCAATGAC 2955

Db     3145 CGAGGAGTTCTCGGAGCTGTACACGTTCAAGTGGACCACCCTGCTGATACCCCCGACCAC 3204

Qy     2956 TCTACTCATCATAAACGTCATTGGAGTCATAGTCGGAGTCTCTGATGCCATCAGCAATGG 3015

Db     3205 GCTCCTCCTGCTGAACTTCATCGGGGTGGTGGCCGGGATCTCGAACGCGATCAACAACGG 3264

Qy     3016 ATACGACTCGTGGGGACCGCTTTTCGGAAGACTGTTCTTTGCACTTTGGGTCATCATTCA 3075

Db     3265 GTACGAGTCGTGGGGCCCCCTGTTCGGGAAGCTCTTCTTCGCCTTCTGGGTGATCGTCCA 3324

Qy     3076 TCTTTACCCGTTCCTTAAAGGTTTGCTTGGGAAACAAGATAGAATGCCAACCATTATTGT 3135

Db     3325 CCTGTACCCGTTCCTCAAGGGTCTGGTGGGGAGGCAGAACAGGACGCCGACGATCGTCAT 3384

Qy     3136 CGTCTGGTCCATCCTCCTGGCCTCGATTCTTACACTTCTTTGGGTCCGGGTTAATCCGTT 3195

Db     3385 CGTCTGGTCCATCCTGCTGGCCTCGATCTTCTCGCTCCTGTGGGTCCGCGTCGACCCGTT 3444

Qy     3196 TGTGGCGAAAGG---CGGTCCTATTCTCGAGATCTGTGGTTTAGACTGCTTGTGATTCGA 3252

Db     3445 CCTCGCCAAGAGCAACGGCCCGCTCCTGGAGGAGTGTGGCCTGGACTGCAACTGAAGTGG 3504

Qy     3253 TTGACCGGTG 3262

Db     3505 GGGCCCCCTG 3514
```

```
RESULT 13
AAZ99524
ID   AAZ99524 standard; DNA; 3813 BP.
XX
AC   AAZ99524;
XX
DT   03-JUL-2000  (first entry)
XX
DE   DNA encoding a maize cellulose synthase.
XX
KW   Maize; cellulose synthase; stalk quality; stand; silage; cellulose;
KW   transgenic plant; plant breeding marker; ss.
XX
OS   Zea mays.
XX
FH   Key             Location/Qualifiers
FT   CDS             215..3493
FT                   /*tag= a
FT                   /product= "cellulose synthase"
FT                   /note= "no termination codon given"
XX
PN   WO200009706-A2.
XX
PD   24-FEB-2000.
XX
PF   16-AUG-1999;   99WO-US18760.
XX
```

PR 17-AUG-1998; 98US-0096822.
XX
PA (PION-) PIONEER HI-BRED INT INC.
XX
PI Dhugga KS, Helentjaris TG, Bowen BA, Wang X;
XX
DR WPI; 2000-224343/19.
DR P-PSDB; AAY84118.
XX
PT New genes which encode maize cellulose synthase polypeptides in plants
PT useful for modulating the expression of cellulose synthase in plants
PT and to produce transgenic plants expressing the novel protein -
XX
PS Claim 1; Page 168-173; 119pp; English.
XX
CC The present sequence encodes a maize cellulose synthase polypeptide.
CC The cellulose synthase can be used for the improvement of stalk quality
CC for improved stand or silage. It also provides an increased concentration
CC of cellulose in the pericarp, hardening the kernel and improving its
CC handling ability. The sequences are used to produce transgenic plants
CC and seeds expressing the cellulose synthase. The polynucleotide is
CC used for modulating, preferably increasing, the level of the synthase
CC in a plant cell. The plants are preferably monocots. The polynucleotide
CC is also used as a probe or primer in the detection quantitation or
CC isolation of gene transcripts. The probes are useful in detecting
CC deficiencies in the level of mRNA in screenings for desired transgenic
CC plant, for detecting mutations in the gene, for monitoring upregulation
CC of expression or changes in enzyme activity in screening assays of
CC compounds, for detection of any number of allelic variants of the gene,
CC or for use as molecular markers in plant breeding programs. The
CC isolated nucleic acids of the present invention can also be used for
CC recombinant expression of their encoded polypeptides or for use as
CC immunogens in the preparation and/or screening of antibodies. The
CC proteins can be employed in assays for enzyme agonists or antagonists
CC of enzyme function or for use of immunogens or antigens to obtain
CC antibodies specifically immunoreactive with a protein.
XX
SQ Sequence 3813 BP; 873 A; 967 C; 1111 G; 862 T; 0 other;

Query Match 41.1%; Score 1416.4; DB 21; Length 3813;
Best Local Similarity 66.7%; Pred. No. 0;
Matches 2128; Conservative 0; Mismatches 1026; Indels 36; Gaps 6;

```
Qy        103 GACATGTCAAATCTGCAGAGATGAGATCGAATTGACTGTTGATGGAGAACCGTTTGTGGC 162
Db        331 GCCGTGCCAGATATGCGGCGACGAGGTCGGGGTGGGCTTCGACGGGGAGCCCTTCGTGGC 390

Qy        163 ATGTAACGAATGTGCATTCCCTGTGTGTAGACCTTGCTATGATACGAAAGACGAGAAGG 222
Db        391 GTGCAACGAGTGCGCCTTCCCGGTCTGCCGCGCCTGCTACGAGTACGAGCGCCGCGAGGG 450

Qy        223 CAATCAAGCTTGTCCACAGTGCAAAACCCGTTTCAAACGTCTTAAAGGAAGTCCAAGAGT 282
Db        451 CTCGCAAGCGTGCCCGCAGTGCAGGACCCGCTACAAGCGCCTCAAGGGCTGCCCGCGGGT 510

Qy        283 TGAAGGTGATGAAGAGGAAGATGACATTGATGATTTAGACAATGAGTT---------TGA 333
Db        511 GGCCGGCGACGAGGAGGAGGACGGCGTCGACGACCTGGAGGGCGAGTTCGGCCTGCAGGA 570

Qy        334 GTATGGAAATAATGGGATTGGATTTGATCAGGTTTCTGAAGGTATGTCAATCTCTCGTCG 393
Db        571 CGGCGCCGCCCACGAGGACGACCCGCAGTACGTCGCCGAGTCCATGCTCAGGGCGCAGAT 630

Qy        394 CAACTCCGGTTTCCCACAATCTGATTTGGATTCAGCTCCACCTGGCTCTCAGATTCCATT 453
Db        631 GAGCTACGGCCGCGGCGGCGACGCGCACCCCGGCTTCAGCCCCGTCCCCAACGTGCCGCT 690

Qy        454 GCTGACTTACGGCGACGAGGACGTTGAGATTTCTTCTGATAGACATGCTCTTATTGTTCC 513
Db        691 CCTCACCAACGGCCAGATGGTTGATGACATCCCGCCGGAGCAGCACGCGCTCGTGCCGTC 750

Qy        514 TCCTTCACTTGGTGGTCAT------GGCAATAGAGTTCATCCTGTTTCTCTTTCTGACCC 567
Db        751 CTACATGAGCGGCGGCGGCGGGGGCAAGAGGATCCACCCGCTCCCTTTCGCAGATCC 810

Qy        568 GACCGTGGCTGCACATCGAAGGCTGATGGTACCTCAGAAAGATCTTGCGGTTTATGGTTA 627
Db        811 CAACCTTCCAGTGCAACCGAGATCCATGGACCCGTCCAAGGATCTGGCCGCCTACGGATA 870

Qy        628 TGGAAGTGTCGCTTGGAAAGATCGGATGGAGGAATGGAAGAGAAAGCAGAATGAGAAACT 687
Db        871 TGGCAGCGTGGCCTGGAAGGAGAGAATGGAGGGCTGGAAGCAGAAGCAG---GAGCGCCT 927

Qy        688 TCAGGTTGTTAGGCATGAAGGAGATCCTGATTTTGAAGATGGTGATGATGCTGATTTTCC 747
Db        928 GCAGCATGTCAGGAGCGAGGGTGGCGGTGATTGG---GATGGCGACGATGCAGATCTGCC 984

Qy        748 AATGATGGATGAGGGAAGGCAGCCATTGTCTATGAAGATACCAATCAAATCGAGCAAGAT 807
Db        985 ACTAATGGATGAAGCTAGGCAGCCATTGTCCAGAAAAGTCCCTATATCATCAAGCCGAAT 1044

Qy        808 AAATCCTTACCGGATGTTAATTGTGCTACGTCTTGTGATTCTTGGTCTCTTCTTTCACTA 867
Db        1045 TAATCCCTACAGGATGATTATCGTTATCCGGTTGGTGGTTTTGGGTTTCTTCTTCCACTA 1104

Qy        868 CCGTATTCTTCACCCCGTCAAAGATGCATATGCTTTGTGGCTTATTTCTGTTATATGTGA 927
Db        1105 CCGAGTGATGCATCCGGCGAAAGATGCATTTGCATTGTGGCTCATATCTGTAATCTGTGA 1164

Qy        928 GATATGGTTTGCTGTTTCATGGGTTCTTGATCAGTTCCCTAAATGGTACCCTATCGAGCG 987
Db        1165 AATCTGGTTTGCGATGTCCTGGATTCTTGATCAGTTCCCAAAGTGGCTTCCAATCGAGAG 1224

Qy        988 AGAAACGTACTTGGACCGACTCTCATTAAGATATGAGAAAGAAGGGAAACCGTCGGGACT 1047
Db        1225 AGAGACTTACCTGGACCGTTTGTCACTAAGGTTTGACAAGGAAGGTCAACCCTCTCAGCT 1284

Qy        1048 ATCCCCTGTGGATGTATTTGTTAGTACAGTGGATCCATTGAAAGAGCCTCCGCTTATTAC 1107
Db        1285 TGCTCCAATCGACTTCTTTGTCAGTACGGTTGATCCCACAAAGGAACCTCCCTTGGTCAC 1344

Qy        1108 TGCAAATACTGTCTTGTCTATTCTTGCTGTTGATTATCCTGTCGATAAGGTTGCTTGTTA 1167
Db        1345 AGCGAACACTGTCCTTTCCATCCTTTCTGTGGATTATCCGGTTGAGAAGGTCTCCTGCTA 1404

Qy        1168 CGTATCTGATGATGGTGCTGCTATGCTTACTTTCGAAGCTCTTTCTGAGACCGCTGAATT 1227
Db        1405 TGTTTCTGATGATGGTGCTGCAATGCTTACGTTTGAAGCATTGTCTGAAACATCTGAATT 1464

Qy        1228 CGCAAGGAAATGGGTTCCTTTCTGCAAGAAATATTGTATTGAGCCTCGTGCTCCCGAATG 1287
Db        1465 TGCAAAGAAATGGGTTCCTTTCAGCAAAAAGTTTAATATCGAGCCTCGTGCTCCTGAGTG 1524

Qy        1288 GTATTTCTGCCATAAAATGGACTACTTGAAGAATAAAGTTCATCCCGCATTTGTTAGGGA 1347
Db        1525 GTACTTCCAACAGAAGATAGACTACCTGAAAGACAAGGTTGCTGCTTCATTTGTTAGGGA 1584

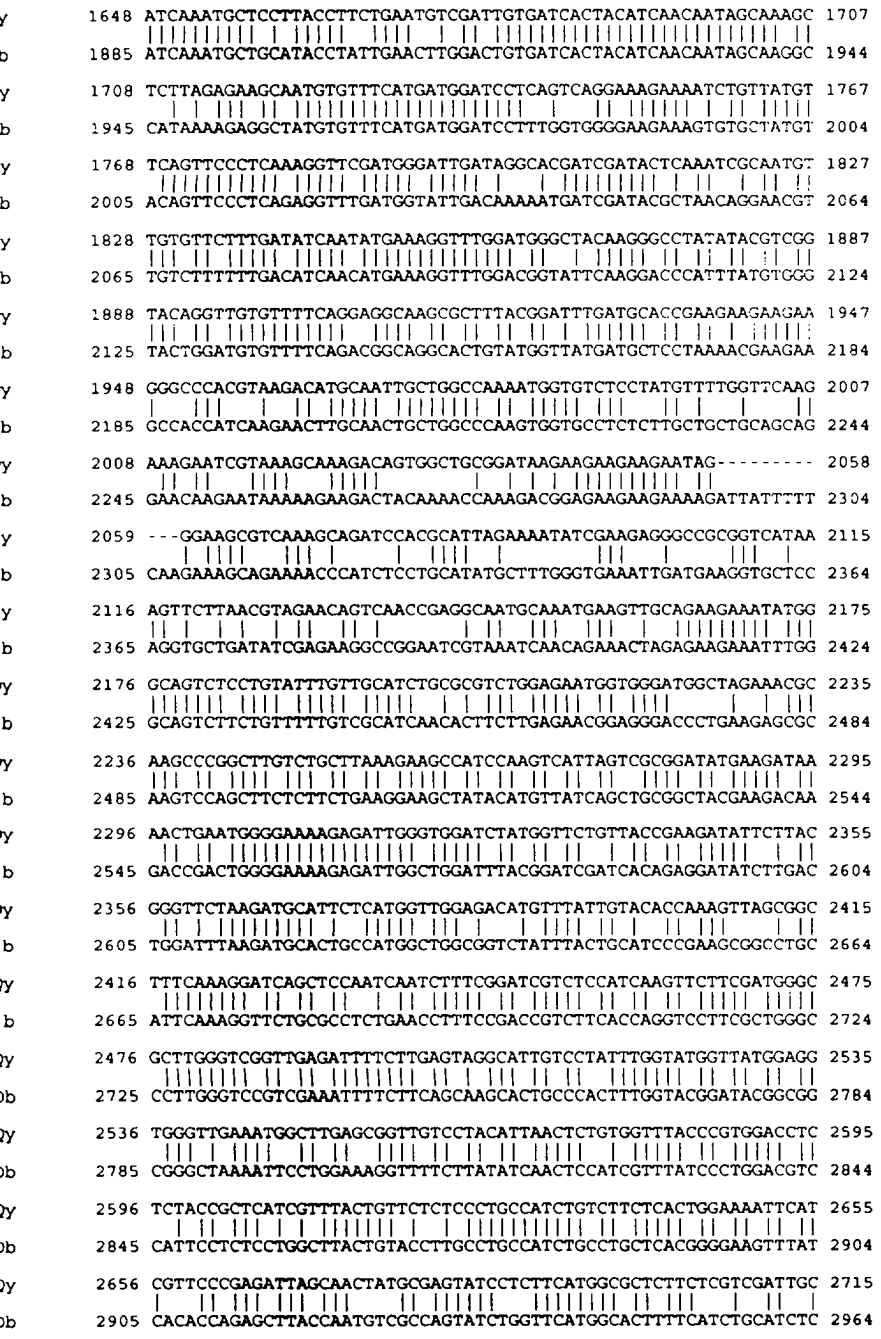
Qy        1348 GCGGCGAGCCATGAAGAGAGATTATGAAGAATTCAAAGTAAAGATCAATGCTTTAGTAGC 1407
Db        1585 GAGGAGGGCGATGAAGAGAGAATACGAGGAATTCAAGGTAAGGATCAATGCCTTGGTTGC 1644

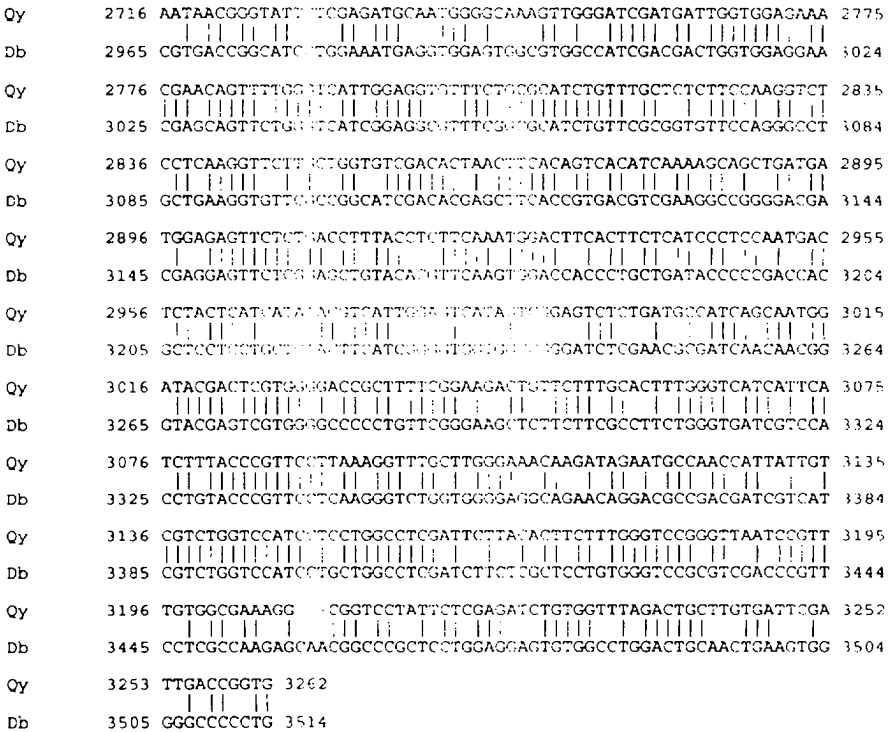
Qy        1408 AACAGCACAGAAAGTGCCTGAGGATGGTTGGACTATGCAAGACGGTACACCTTGGCCCGG 1467
Db        1645 AAAAGCCCAAAAGGTTCCTGAGGAAGGATGGACAATGCAAGATGGAAGCCCCTGGCCTGG 1704

Qy        1468 TAATAGTGTGCGAGATCATCCTGGCATGATTCAGGTCTTCCTTGGAAGTGACGGTGTTCG 1527
Db        1705 AAACAACGTACGCGATCATCCTGGAATGATTCAGGTATTCCTTGGCCAAAGTGGCGGTCG 1764

Qy        1528 TGATGTCGAAAACAACGAGTTGCCTCGATTAGTTTACGTTTCTCGTGAGAAGAGACCCGG 1587
Db        1765 TGATGTGGAAGGAAATGAGTTGCCTCGCCTGGTTTATGTCTCGAGAGAAAAGAGGCCAGG 1824

Qy        1588 ATTTGATCACCATAAGAAGGCTGGAGCTATGAATTCCCTGATACGAGTCTCTGGGGTTCT 1647
Db        1825 TTATAACCATCACAAGAAGGCTGGTGCCATGAATGCACTGGTCCGTGTCTCTGCTGTCTT 1884
```





```
RESULT 14
AAV06565
ID   AAV06565 standard; cDNA; 3603 BP.
XX
AC   AAV06565;
XX
DT   06-JUL-1998  (first entry)
XX
DE   Arabidopsis cellulose synthase RSW1 cDNA clone.
XX
KW   Beta-glycosyltransferase; cellulose synthase; RSW1 gene;
KW   beta-1,4-glucan; transgenic plant; ss.
XX
OS   Arabidopsis thaliana var. Columbia.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..3246
FT                   /*tag= a
XX
PN   WO9800549-A1.
XX
PD   08-JAN-1998.
XX
PF   24-JUN-1997;   97WO-AU00402.
XX
PR   27-JUN-1996;   96AU-0000699.
XX
PA   (CSIR ) COMMONWEALTH SCI & IND RES ORG.
PA   (AUSU ) UNIV AUSTRALIAN NAT.
XX
PI   Arioli A,  Betzner AS,  Peng L,  Williamson RE;
XX
```

DR WPI; 1998-086974/08.
DR P-PSDB; AAW33817.
XX
PT DNA encoding cellulose biosynthetic enzyme - useful for manipulation
PT of cellulose and beta-1,4-glucan
XX
PS Claim 12; Page 109-116; 207pp; English.
XX
CC This cDNA molecule codes for the RSW1 cellulose synthase (see
CC AAW33817) of Arabidopsis thaliana. The sequence is derived from 2
CC cDNA clones isolated from a cDNA library using probes obtained by
CC PCR amplification of Arabidopsis cDNA with primers (see AAT99632-34)
CC based on RSW1 genomic clone 23H12 (see AAV06563) and EST clone AAT20782
CC (see AAV06562). Claimed nucleic acid molecules (see AAV06562-69)
CC coding for claimed polypeptides (see AAW33816-20 and AAW46202) involved
CC in cellulose biosynthesis can be used to manipulate the cellulose
CC and/or beta-glucan content of transgenic plants. Expression of
CC nucleic acids in the sense orientation increases the level of
CC cellulose and reduces the level of non-crystalline beta-1,4-glucan
CC and starch, providing plants with modified strength and/or shape
CC and/or fibre properties, or having increased resistance to stresses
CC or pests. Antisense, ribozyme or co-suppression molecules can be
CC used to reduce the cellulose content of a transgenic plant, e.g. to
CC improve digestibility or to alter carbon partitioning such that
CC increased carbon is available for growth, rather than deposited as
CC cellulose.
XX
SQ Sequence 3603 BP; 950 A; 745 C; 852 G; 1056 T; 0 other;

Query Match 37.6%; Score 1294.8; DB 19; Length 3603;
Best Local Similarity 64.5%; Pred. No. 0;
Matches 2074; Conservative 0; Mismatches 1082; Indels 60; Gaps 7;

```
Qy          9 GGTTAATCGCCGGTTCTCACAACAGGAATGAGTTTGTCCTCATTAATGCCGATGAGAATG 68
              | ||  | ||  || ||  ||   || || ||| | || |  ||        ||     |||
Db         17 GCTTGGTTGCTGGATCCTACCGGAGAAACGAGCTCGTTCGGATCCGACATGAATCTGATG 76

Qy         69 CCCGAATAAGATCAGTCCAAGAGCTGAGTGGACAGACATGTCAAATCTGCAGAGATGAGA 128
              | | |  | | |  |  | :  ||| ||| |||| |||||| |||||  | ||||||
Db         77 GCGGGACCAAACCTTTGAAGAATATGAATGGCCAGATATGTCAGATCTGTGGTGATGATG 136

Qy        129 TCGAATTGACTGTTGATGGAGAACCGTTTGTGGCATGTAACGAATGTGCATTCCCTGTGT 188
              | | | |  |||    |||||| ||||| || ||||| |||||||| ||||||||||
Db        137 TTGGACTCGCTGAAACTGGAGATCTCTTTGTCGCGTGTAATGAATGTGCCTTCCCTGTGT 196

Qy        189 GTAGACCTTGCTATGAGTACGAAAGACGAGAAGGCAATCAAGCTTGTCCACAGTGCAAAA 248
              || | ||||||||||||||||| .|   ||, || | |||   ||| || || ||||| |
Db        197 GTCGGCCTTGCTATGAGTACGAGAGGAAAGATGGAACTCAGTGTTGCCCTCAATGCAAGA 256

Qy        249 CCCGTTTCAAACGTCTTAAAGGAATCTGTAAAAGTTGAAGG-ATGAAGAGGAAGATGACA 308
              | . |||| ||| | |  || .|   ||| ||| |||.| || ||||
Db        257 CTAGATTCAGACGACACAGGGGGAGTCTCTGTGTTGAAGGAATGAAGATGAGGATGATG 316

Qy        309 TTGATGATTTAGACAATGAGTTTAAGTATGGAAATAATGGAATTGGATTTGATCAGGTTT 368
              |||||||| | || ||||||||  | ||                        | ||||
Db        317 TTGATGATATCGAGAATGAGTTCAATTA--  - -----      --- CGCCCAGGGAG 355

Qy        369 CTGAAGGTATGTCAATCTCTCGTCGCAACTCCGGTTTCCCAGAATCTGATTTGGATTCAG 428
              || |     | | |  || :   | | ||  |   ||  |   || | |
Db        356 CTAACAAGGCGAGACACCAACGCCATGGCGAAGAGTTTTCTTCTTCCTCTAGACATGAAT 415

Qy        429 CTCCACCTGGCTCTCAGATTCCATTGCTGACTTACGGCGACGAGGACGTTGAGATTTCTT 488
              ||| |||     |||  | |  |  |  |  |  || ||  |  ||
Db        416 CTCAACCAATTCCTCTTCTCACCCATGGCCATACGGTTTCTGGAGAGATTCGCACGCCTG 475

Qy        489 CTGATAGACATGCTCTTATTGTTCCTCCTTCACTTGGTGGTCATGGCAATAGAGTTCATC 548
              |     | ||  |  | |    |  | |  : ||| |  || ||  |  | |  |
Db        476 ATACACAATCTGTGCGAACTACATCAGGTCCTTTGGGTCCTTCTGACAGGAATGCTATTT 535

Qy        549 CTGTTTCTCTTTCTGACCCGACCGTGGCTGCACATCGAAGGCTGATGGTACCTCAGAAAG 608
              |  | |   | ||| ||        | |  : |  ||| :  ||| ||    ||||
Db        536 CATCTCCATATATTGATCCACGGCAACCTGTCCCTGTAAGAATCGTGGACCCGTCAAAAG 595

Qy        609 ATCTTGCGGTTTATGGTTATGGAAGTGTCGCTTGGAAAGATCGGATGGAGGAATGGAAGA 668
              | |      ||||||  ||| | ||| |  |||||||  |  | || |  ||||||
Db        596 ACTTGAACTCTTATGGGCTTGGTAATGTTGACTGGAAAGAAAGAGTTGAAGGCTGGAAGC 655

Qy        669 GAAAGCAGAATGAGAAACT----------------TCAGGTTGTTAGGCATGAAGGAG 710
                ||||||  |  | ||  |                |  ||  || | |  |||||
Db        656 TGAAGCAGGAGAAAAATATGTTACAGATGACTGGTAAATACCATGAAGGGAAAGGAGGAG 715

Qy        711 ATCCTGATTTTGAAGATGGTGATGATGCTGATTTTCCAATGATGGATGAGGGAAGGCAGC 770
              |   |||        | |   ||| |  ||  | | ||||  |||||   | | |  |
Db        716 AAATTGAAGGGACTGGTTCCAATGGCGAAGAACTCCAAATGGCTGATGATACACGTCTTC 775

Qy        771 CATTGTCTATGAAGATACCAATCAAATCGAGCAAGATAAATCCTTACCGGATGTTAATTG 830
              |  ||  |      | | || |||  |||         |||  ||||| ||| |  | |||
Db        776 CTATGAGTCGTGTGGTGCCTATCCCATCTTCTCGCCTAACCCCTTATCGGGTTGTGATTA 835

Qy        831 TGCTACGTCTTGTGATTCTTGGTCTCTTCTTTCACTACCGTATTCTTCACCCCGTCAAAG 890
              | || || ||| | ||  |  || ||||||| || || ||||     ||||||  || |||
Db        836 TTCTCCGGCTTATCATCTTGTGTTTCTTCTTGCAATATCGTACAACTCACCCTGTGAAAA 895

Qy        891 ATGCATATGCTTTGTGGCTTATTTCTGTTATATGTGAGATATGGTTTGCTGTTTCATGGG 950
              |||||||| |||||||| | |  || ||||| |||||||| ||||||||  |||| |||
Db        896 ATGCATATCCTTTGTGGTTGACCTCGGTTATCTGTGAGATCTGGTTTGCATTTTCTTGGC 955

Qy        951 TTCTTGATCAGTTCCCTAAATGGTACCCTATCGAGCGAGAAACGTACTTGGACCGACTCT 1010
              |||||||||||||| || ||||||||||| ||  |  | || || ||  | ||||| |||
Db        956 TTCTTGATCAGTTTCCCAAATGGTACCCCATTAACAGGGAGACTTATCTTGACCGTCTCG 1015

Qy       1011 CATTAAGATATGAGAAAGAAGGGAAACCGTCGGGACTATCCCCTGTGGATGTATTTGTTA 1070
              |  ||||||||||  ||| ||  |||| ||     ||||  ||||| |||||  |||||||
Db       1016 CTATAAGATATGATCGAGACGGTGAACCATCACAGCTCGTTCCTGTTGATGTGTTTGTTA 1075

Qy       1071 GTACAGTGGATCCATTGAAAGAGCCTCCGCTTATTACTGCAAATACTGTCTTGTCTATTC 1130
              |||||||||| |||||||||||||||| ||| |||| |||||  || ||  | || ||||
Db       1076 GTACAGTGGACCCATTGAAAGAGCCTCCCCTTGTTACAGCAAACACAGTTCTCTCGATTC 1135

Qy       1131 TTGCTGTTGATTATCCTGTCGATAAGGTTGCTTGTTACGTATCTGATGATGGTGCTGCTA 1190
              || |||| || || || || ||||| || || ||||| || || ||||||||| | ||||
Db       1136 TTTCTGTGGACTACCCGGTAGATAAAGTAGCCTGTTATGTTTCAGATGATGGTTCAGCTA 1195

Qy       1191 TGCTTACTTTCGAAGCTCTTTCTGAGACCGCTGAATTCGCAAGGAAATGGGTTCCTTTCT 1250
              ||||||| || ||| | |||||||| |||||||| || |||| |||||||| || || |
Db       1196 TGCTTACCTTTGAATCCCTTTCTGAAACCGCTGAGTTTGCAAAGAAATGGGTACCATTTT 1255

Qy       1251 GCAAGAAATATTGTATTGAGCCTCGTGCTCCCGAATGGTATTTCTGCCATAAAATGGACT 1310
              |||||||||     ||||| ||| | || || |||| |||||   ||| || || | |
Db       1256 GCAAGAAATTCAACATTGAACCTAGGGCCCCTGAATTCTATTTTGCCCAGAAGATAGATT 1315

Qy       1311 ACTTGAAGAATAAAGTTCATCCCGCATTTGTTAGGGAGCGGCGAGCCATGAAGAGAGATT 1370
              |||||||| | ||  | || ||  | |||||||  |||||  ||||| |||||||||||| |
Db       1316 ACTTGAAGGACAAGATCCAACCGTCTTTTGTTAAAGAGCGACGAGCTATGAAGAGAGAGT 1375

Qy       1371 ATGAAGAATTCAAAGTAAAGATCAATGCTTTAGTAGCAACAGCACAGAAAGTGCCTGAGG 1430
              ||||||| || ||||| | ||| ||||||  | || || | |||||||||| | |||||  |
Db       1376 ATGAAGAGTTTAAAGTGAGGATAAATGCTCTTGTTGCCAAAGCACAGAAAATCCCTGAAG 1435

Qy       1431 ATGGTTGGACTATGCAAGACGGTACACCTTGGCCCGGTAATAGTGTGCGAGATCATCCTG 1490
              | || ||||| ||||| || ||||| || ||||| |||||  |      ||||||||||
Db       1436 AAGGCTGGACAATGCAGGATGGTACTCCCTGGCCTGGTAACAACACTAGAGATCATCCTG 1495

Qy       1491 GCATGATTCAGGTCTTCCTTGGAAGTGACGGTGTTCGTGATGTCGAAAACAACGAGTTGC 1550
              | ||||| ||||| ||| | ||    |  || | ||  |||  |||     || ||| |||
Db       1496 GAATGATACAGGTGTTCTTAGGCCATAGTGGGGGTCTGGATACCGATGGAAATGAGCTGC 1555

Qy       1551 CTCGATTAGTTTACGTTTCTCGTGAGAAGAGACCCGGATTTGATCACCATAAGAAGGCTG 1610
              || || |  | || |||||||||| ||| | || |||||| | ||||| || |||||||
Db       1556 CTAGACTCATCTATGTTTCTCGTGAAAAGCGGCCTGGATTTCAACACCACAAAAAGGCTG 1615

Qy       1611 GAGCTATGAATTCCCTGATACGAGTCTCTGGGGTTCTATCAAATGCTCCTTACCTTCTGA 1670
              |||||||||| | |||| || || |||| |||||  | ||||   | || ||| |||
Db       1616 GAGCTATGAATGCATTGATCCGTGTATCTGCTGTTCTTACCAATGGAGCATATCTTTTGA 1675

Qy       1671 ATGTCGATTGTGATCACTACATCAACAATAGCAAAGCTCTTAGAGAAGCAATGTGTTTCA 1730
```

```
b      1676 ACGTGGATTGTGATCATTACTTTAATAACAGTAAGGCTATTAAAGAAGCTATGTGTTTCA 1735
y      1731 TGATGGATCCTCAGTCAGGAAAGAAAATCTGTTATGTTCAGTTCCCTCAAAGGTTCGATG 1790
b      1736 TGATGGACCCGGCTATTGGAAAGAAGTGCTGCTATGTCCAGTTCCCTCAACGTTTTGACG 1795
y      1791 GGATTGATAGGCACGATCGATACTCAAATCGCAATGTTGTGTTCTTTGATATCAATATGA 1850
b      1796 GTATTGATTTGCACGATCGATATGCCAACAGGAATATAGTCTTTTTCGATATTAACATGA 1855
y      1851 AAGGTTTGGATGGGCTACAAGGGCCTATATACGTCGGTACAGGTTGTGTTTTCAGGAGGC 1910
b      1856 AGGGGTTGGATGGTATCCAGGGTCCAGTATATGTGGGTACTGGTTGTTGTTTTAATAGGC 1915
y      1911 AAGCGCTTTACGGATTTGATGCACCGAAGAAGAAGAAGGGCCCACGTAAGACATGCAATT 1970
b      1916 AGGCTCTATATGGGTATGATCCTGTTTTGACGGAAGAAG----ATTTAGAACCAAATATT 1971
y      1971 GCTGGCCAAAAATGGTGTCTCCTATGTTTTGGTTCAAGAAAGAATCGTAAAGCAAAGACAG 2030
b      1972 ATTGTCAAGA-----------GCTGTTGCGGGTCAAGGAAGAAAGGTAAAAGTAGCAAGA 2020
y      2031 TGGCTGCGGATAAGAAGA--AGAAGAATAGGGAAGCGTCAAAGCAGATCCACGCATTAGA 2088
b      2021 AGTATAACTACGAAAAGAGGAGAGGCATCAACAGAAGTGACTCCAATGCTCCACTTTTCA 2080
y      2089 AAATATCGAAGAGGGCCG-CGGTCATAAAGTTCTTAACGTAGAACAGTCAACCGAGGCAA 2147
b      2081 ATATGGAGGACATCGATGAGGGTTTTGAAGGTTATGATGATGAGAGGTCTATTCTAATGT 2140
y      2148 TGCAAATGAAGTTGCAGAAGAAATATGGGCAGTCTCCTGTATTTGTTGCATCTGCGCGTC 2207
b      2141 CCCAGAGGAGTGTAGAGAAGCGTTTTGGTCAGTCGCCGGTATTTATTGCGGCAACCTTCA 2200
y      2208 TGGAGAATGGTGGGATGGCTAGAAACGCAAGCCCGGCTTGTCTGCTTAAAGAAGCCATCC 2267
b      2201 TGGAACAAGGCGGCATTCCACCAACAACCAATCCCGCTACTCTTCTGAAGGAGGCTATTC 2260
y      2268 AAGTCATTAGTCGCGGATATGAAGATAAAACTGAATGGGGAAAAGAGATTGGGTGGATCT 2327
b      2261 ATGTTATAAGCTGTGGTTACGAAGACAAGACTGAATGGGGCAAAGAGATTGGTTGGATCT 2320
y      2328 ATGGTTCTGTTACCGAAGATATTCTTACGGGTTCTAAGATGCATTCTCATGGTTGGAGAC 2387
b      2321 ATGGTTCCGTGACGGAAGATATTCTTACTGGGTTCAAGATGCATGCCCGGGGTTGGATAT 2380
y      2388 ATGTTTATTGTACACCAAAGTTAGCGGCTTTCAAAGGATCAGCTCCAATCAATCTTTCGG 2447
b      2381 CGATCTACTGCAATCCTCCACGCCCTGCGTTCAAGGGATCTGCACCAATCAATCTTTCTG 2440
y      2448 ATCGTCTCCATCAAGTTCTTCGATGGGCGCTTGGGTCGGTTGAGATTTTCTTGAGTAGGC 2507
b      2441 ATCGTTTGAACCAAGTTCTTCGATGGGCTTTGGGATCTATCGAGATTCTTCTTAGCAGAC 2500
y      2508 ATTGTCCTATTTGGTATGGTTATGGAGGTGGGTTGAAATGGCTTGAGCGGTTGTCCTACA 2567
b      2501 ATTGTCCTATCTGGTATGGTTACCATGGAAGGTTGAGACTTTTGGAGAGGATCGCTTATA 2560
y      2568 TTAACTCTGTGGTTTACCCGTGGACCTCTCTACCGCTCATCGTTTACTGTTCTCTCCCTG 2627
b      2561 TCAACACCATCGTCTATCCTATTACATCCATCCCTCTTATTGCGTATTGTATTCTTCCCG 2620
y      2628 CCATCTGTCTTCTCACTGGAAAATTCATCGTTCCCGAGATTAGCAACTATGCGAGTATCC 2687
b      2621 CTTTTTGTCTCATCACCGACAGATTCATCATACCCGAGATAAGCAACTACGCGAGTATTT 2680
y      2688 TCTTCATGGCGCTCTTCTCGTCGATTGCAATAACGGGTATTCTCGAGATGCAATGGGGCA 2747
b      2681 GGTTCATTCTACTCTTCATCTCAATTGCTGTGACTGGAATCCTGGAGCTGAGATGGAGCG 2740
y      2748 AAGTTGGGATCGATGATTGGTGGAGAAACGAACAGTTTTGGGTCATTGGAGGTGTTTCTG 2807
```

```
Db     2741 GTGTGAGCATTGAAGATTGGTGGAGGAACGAGCAGTTCTGGGTCATTGGTGGCACATCCG 2800
Qy     2808 CGCATCTGTTTGCTGTCTTCCAAGGTCTCCTCAAGGTTCTTGCTGGTGTCGACACTAACT 2867
Db     2801 CCCATCTTTTTGCTGTCTTCCAAGGTCTACTTAAGGTTCTTGCTGGTATCGACACCAACT 2860
Qy     2868 TCACAGTCACATCAAAAG---CAGCTGATGATGGAGAGTTCTCTGACCTTTACCTCTTCA 2924
Db     2861 TCACCGTTACATCTAAAGCCACAGACGAAGATGGGGATTTTGCAGAACTCTACATCTTCA 2920
Qy     2925 AATGGACTTCACTTCTCATCCCTCCAATGACTCTACTCATCATAAACGTCATTGGAGTCA 2984
Db     2921 AATGGACAGCTCTTCTCATTCCACCAACCACCGTCCTACTTGTGAACCTCATAGGCATTG 2980
Qy     2985 TAGTCGGAGTCTCTGATGCCATCAGCAATGGATACGACTCGTGGGGACCGCTTTTCGGAA 3044
Db     2981 TGGCTGGTGTCTCTATGCTGTAAACAGTGGTACCAGTCGTGGGGTCCGCTTTTCGGGA 3040
Qy     3045 GACTGTTCTTTGCACTTTGGGTCATCATTCATCGTTTACCCGTTCCTTAAAGGTTTGCTTG 3104
Db     3041 AGCTCTTCTTCGCCTTATGGGTTATTGCCCATCTCTACCCTTTCTTGAAAGGTCTGTTGG 3100
Qy     3105 GGAAACAAGATAGAATGCCAACCATTATTGTCGTCTGGTCCATCCTCCTGGCCTCGATTC 3164
Db     3101 GAAGACAAAACCGAACACCAACCATCGTCATTGTCTGGTCTGTTCTTCTCGCCTCCATCT 3160
Qy     3165 TTACACTTCTTTGGGTCCGGGTTAATCCGTTTGTGG 3200
Db     3161 TCTCGTTGCTTTGGGTCAGGATCAATCCCTTTGTGG 3196
```

RESULT 15
AAV06568
ID AAV06568 standard; cDNA; 3673 BP.
XX
AC AAV06568;
XX
DT 06-JUL-1998 (first entry)
XX
DE Arabidopsis cellulose synthase mutant rsw1 cDNA.
XX
KW Beta-glycosyltransferase; cellulose synthase; rsw1 gene;
KW beta-1,4-glucan; transgenic plant; ss
XX
OS Arabidopsis thaliana var. Columbia rsw1 mutant.
XX
FH Key Location/Qualifiers
FT CDS 71..3316
FT /*tag= a
FT mutation 1716
FT /*tag= b
FT /note= "codon GTT (Val) is GCT (Ala) in
FT wild-type RSW1 gene"
XX
PN WO9800549-A1.
XX
PD 08-JAN-1998.
XX
PF 24-JUN-1997; 97WO-AU00402.
XX
PR 27-JUN-1996; 96AU-0000699.
XX
PA (CSIR) COMMONWEALTH SCI & IND RES ORG.
PA (AUSU) UNIV AUSTRALIAN NAT.
XX
PI Arioli A, Betzner AS, Peng L, Williamson RE;
XX
DR WPI; 1998-086974/08.
DR P-PSDB; AAW44820.
XX
PT DNA encoding cellulose biosynthetic enzyme - useful for manipulation
PT of cellulose and beta-1,4-glucan
XX

PS Claim 12; Page 149-157; 207pp; English.
XX
CC This cDNA molecule codes for the rsw1 mutant cellulose synthase
CC (see AAW33820) of Arabidopsis thaliana mutant rsw1. It was obtained
CC by PCR amplification (see AAT99635-36). The rsw1 gene is a mutant
CC cellulose gene that produces reduced inflorescence length, reduced
CC fertility, misshapen epidermal cells, reduced cellulose content and
CC the accumulation of non-crystalline beta-1,4-glucan when expressed
CC in plant cells. Wild-type RSW1 and RSW1-like genomic clones and
CC cDNA clones (see AAV06562-67) of Arabidopsis have been isolated and
CC shown to encode putative cellulose synthases (see AAW33816-19).
CC Claimed nucleic acid molecules (see AAV06562-69) coding for claimed
CC polypeptides (see AAW33816-20 and AAW46202) involved in cellulose
CC biosynthesis can be used to manipulate the cellulose and/or
CC beta-glucan content of transgenic plants, thereby modifying
CC strength and/or shape and/or fibre properties, increasing
CC resistance to stresses or pests, improve digestibility or altering
CC carbon partitioning so that increased carbon is available for
CC growth.
XX
SQ Sequence 3673 BP; 961 A; 760 C; 867 G; 1085 T; 0 other;

Query Match 37.6%; Score 1294.8; DB 19; Length 3673;
Best Local Similarity 64.5%; Pred. No. 0;
Matches 2074; Conservative 0; Mismatches 1082; Indels 60; Gaps 7;

```
Qy          9 GGTTAATCGCCGGTTCTCACAACAGGAATGAGTTTGTCCTCATTAATGCCGATGAGAATG 68
              |  ||  | || ||  ||   ||  || ||| | ||  |  ||       ||     |||
Db         87 GCTTGGTTGCTGGATCCTACCGGAGAAACGAGCTCGTTCGGATCCGACATGAATCTGATG 146

Qy         69 CCCGAATAAGATCAGTCCAAGAGCTGAGTGGACAGACATGTCAAATCTGCAGAGATGAGA 128
                | | |  | | |  |  |  | |||  ||| |||| |||||| |||||  | |||||
Db        147 GCGGGACCAAACCTTTGAAGAATATGAATGGCCAGATATGTCAGATCTGTGGTGATGATG 206

Qy        129 TCGAATTGACTGTTGATGGAGAACCGTTTGTGGCATGTAACGAATGTGCATTCCCTGTGT 188
              | | | |  |||     |||||  || ||||| ||||||| ||||||||||
Db        207 TTGGACTCGCTGAAACTGGAGATGTCTTTGTCGCGTGTAATGAATGTGCCTTCCCTGTGT 266

Qy        189 GTAGACCTTGCTATGAGTACGAAAGACGAGAAGGCAATCAAGCTTGTCCACAGTGCAAAA 248
              ||  |||||||||||||||||| ||   ||| || | |||   ||| || || ||||| |
Db        267 GTCGGCCTTGCTATGAGTACGAGAGGAAAGATGGAACTCAGTGTTGCCCTCAATGCAAGA 326

Qy        249 CCCGTTTCAAACGTCTTAAAGGAAGTCCAAGAGTTGAAGGTGATGAAGAGGAAGATGACA 308
              |  | |||| ||| |  |  || |||||  | |||||||| |||||||| || |||||
Db        327 CTAGATTCAGACGACACAGGGGGAGTCCTCGTGTTGAAGGAGATGAAGATGAGGATGATG 386

Qy        309 TTGATGATTTAGACAATGAGTTTGAGTATGGAAATAATGGGATTGGATTTGATCAGGTTT 368
              |||||||| | || ||||||||  | ||                  | ||||
Db        387 TTGATGATATCGAGAATGAGTTCAATTA---- -------      -----CGCCCAGGGAG 425

Qy        369 CTGAAGGTATGTCAATCTCTCGT-GCAACT-CGGTTTCCCA-AATCTGATTTGGATTCAG 428
              ||  |   |  | |  |  |  | |  |  ||  |  |  |  |  |  |  |  |
Db        426 CTAACAAGGCGAGACACCAACGCCATGGCGAAGAGTTTTCTT-TTCCTCTAGACATGAAT 485

Qy        429 CTCCACCTGGCTCTCAGATTCCATTGCTGACTTACGGCGACGAGGACGTTGAGATTTCTT 488
              |||  |||     |||  |  |       |  |     | ||  ||   |  ||
Db        486 CTCAACCAATTCCTCTTCTCACCCATGGCCATACGGTTTCTGGAGAGATTCGCACGCCTG 545

Qy        489 CTGATAGACATGCTCTTATTGTTCCTCCTTCACTTGGTGGTCATGGCAATAGAGTTCATC 548
              |     |  ||  | |  |  | | | ||| |  || ||  | |  | |  |  |
Db        546 ATACACAATCTGTGCGAACTACATCAGGTCCTTTGGGTCCTTCTGACAGGAATGCTATTT 605

Qy        549 CTGTTTCTCTTTCTGACCCGACCGTGGCTGCACATCGAAGGCTGATGGTACCTCAGAAAG 608
              |  | |  |  ||| ||      |||  | |  |||  |  |||  ||    ||||
Db        606 CATCTCCATATATTGATCCACGGCAACCTGTCCCTGTAAGAATCGTGGACCCGTCAAAAG 665

Qy        609 ATCTTGCGGTTTATGGTTATGGAAGTGTCGCTTGGAAAGATCGGATGGAGGAATGGAAGA 668
              |  |      ||||||   ||  | |||  |  ||||||||     |  || |  ||||||
Db        666 ACTTGAACTCTTATGGGCTTGGTAATGTTGACTGGAAAGAAAGAGTTGAAGGCTGGAAGC 725

Qy        669 GAAAGCAGAATGAGAAACT----------------TCAGGTTGTTAGGCATGAAGGAG 710
                 ||||||  |  ||  |                |  ||   ||  |  | |||||
Db        726 TGAAGCAGGAGAAAAATATGTTACAGATGACTGGTAAATACCATGAAGGGAAAGGAGGAG 785

Qy        711 ATCCTGATTTTGAAGATGGTGATGATGCTGATTTTCCAATGATGGATGAGGGAAGGCAGC 770
              |  |||       | |   |||  |  ||  | | ||||  |||||   | | |  |
Db        786 AAATTGAAGGGACTGGTTCCAATGGCGAAGAACTCCAAATGGCTGATGATACACGTCTTC 845

Qy        771 CATTGTCTATGAAGATACCAATCAAATCGAGCAAGATAAATCCTTACCGGATGTTAATTG 830
              |  ||  |     | | ||| |||  |||          |||  ||||| ||| |  | |||
Db        846 CTATGAGTCGTGTGGTGCCTATCCCATCTTCTCGCCTAACCCCTTATCGGGTTGTGATTA 905

Qy        831 TGCTACGTCTTGTGATTCTTGGTCTCTTCTTTCACTACCGTATTCTTCACCCCGTCAAAG 890
              | || || ||| | ||  |  || ||||||| || || ||||    |||||| || |||
Db        906 TTCTCCGGCTTATCATCTTGTGTTTCTTCTTGCAATATCGTACAACTCACCCTGTGAAAA 965

Qy        891 ATGCATATGCTTTGTGGCTTATTTCTGTTATATGTGAGATATGGTTTGCTGTTTCATGGG 950
              |||||||| |||||||| | |  || ||||| |||||||| ||||||||  |||| |||
Db        966 ATGCATATCCTTTGTGGTTGACCTCGGTTATCTGTGAGATCTGGTTTGCATTTTCTTGGC 1025

Qy        951 TTCTTGATCAGTTCCCTAAATGGTACCCTATCGAGCGAGAAACGTACTTGGACCGACTCT 1010
              |||||||||||||| || |||||||||| ||  |  | || ||  | ||||| |||
Db       1026 TTCTTGATCAGTTTCCCAAATGGTACCCCATTAACAGGGAGACTTATCTTGACCGTCTCG 1085

Qy       1011 CATTAAGATATGAGAAAGAAGGGAAACCGTCGGGACTATCCCCTGTGGATGTATTTGTTA 1070
              |  ||||||||||    ||| ||  |||| ||      ||||| ||||| |||||||
Db       1086 CTATAAGATATGATCGAGACGGTGAACCATCACAGCTCGTTCCTGTTGATGTGTTTGTTA 1145

Qy       1071 GTACAGTGGATCCATTGAAAGAGCCTCCGCTTATTACTGCAAATACTGTCTTGTCTATTC 1130
              |||||||||| |||||||||||||||| ||| |||| ||||| || ||  | || ||||
Db       1146 GTACAGTGGACCCATTGAAAGAGCCTCCCCTTGTTACAGCAAACACAGTTCTCTCGATTC 1205

Qy       1131 TTGCTGTTGATTATCCTGTCGATAAGGTTGCTTGTTACGTATCTGATGATGGTGCTGCTA 1190
              || |||| || || || || ||||| || || ||||| || || |||||||||  | ||||
Db       1206 TTTCTGTGGACTACCCGGTAGATAAAGTAGCCTGTTATGTTTCAGATGATGGTTCAGCTA 1265

Qy       1191 TGCTTACTTTCGAAGCTCTTTCTGAGACCGCTGAATTCGCAAGGAAATGGGTTCCTTTCT 1250
              ||||||| || ||| | |||||||| |||||||| || |||| ||||||||| || || |
Db       1266 TGCTTACCTTTGAATCCCTTTCTGAAACCGCTGAGTTTGCAAAGAAATGGGTACCATTTT 1325

Qy       1251 GCAAGAAATATTGTATTGAGCCTCGTGCTCCCGAATGGTATTTCTGCCATAAAATGGACT 1310
              |||||||||     ||||| ||| | || || |||| |||||    ||| || || || |
Db       1326 GCAAGAAATTCAACATTGAACCTAGGGCCCCTGAATTCTATTTTGCCCAGAAGATAGATT 1385

Qy       1311 ACTTGAAGAATAAAGTTCATCCCGCATTTGTTAGGGAGCGGCGAGCCATGAAGAGAGATT 1370
              |||||||| | ||  | || ||  | |||||||  ||||| ||||| ||||||||||| |
Db       1386 ACTTGAAGGACAAGATCCAACCGTCTTTTGTTAAAGAGCGACGAGCTATGAAGAGAGAGT 1445

Qy       1371 ATGAAGAATTCAAAGTAAAGATCAATGCTTTAGTAGCAACAGCACAGAAAGTGCCTGAGG 1430
              ||||||| || ||||| | ||| |||||| | || || | |||||||||| | ||||| |
Db       1446 ATGAAGAGTTTAAAGTGAGGATAAATGCTCTTGTTGCCAAAGCACAGAAAATCCCTGAAG 1505

Qy       1431 ATGGTTGGACTATGCAAGACGGTACACCTTGGCCCGGTAATAGTGTGCGAGATCATCCTG 1490
              | || ||||| ||||| || ||||| || ||||| ||||| |          ||||||||||
Db       1506 AAGGCTGGACAATGCAGGATGGTACTCCCTGGCCTGGTAACAACACTAGAGATCATCCTG 1565

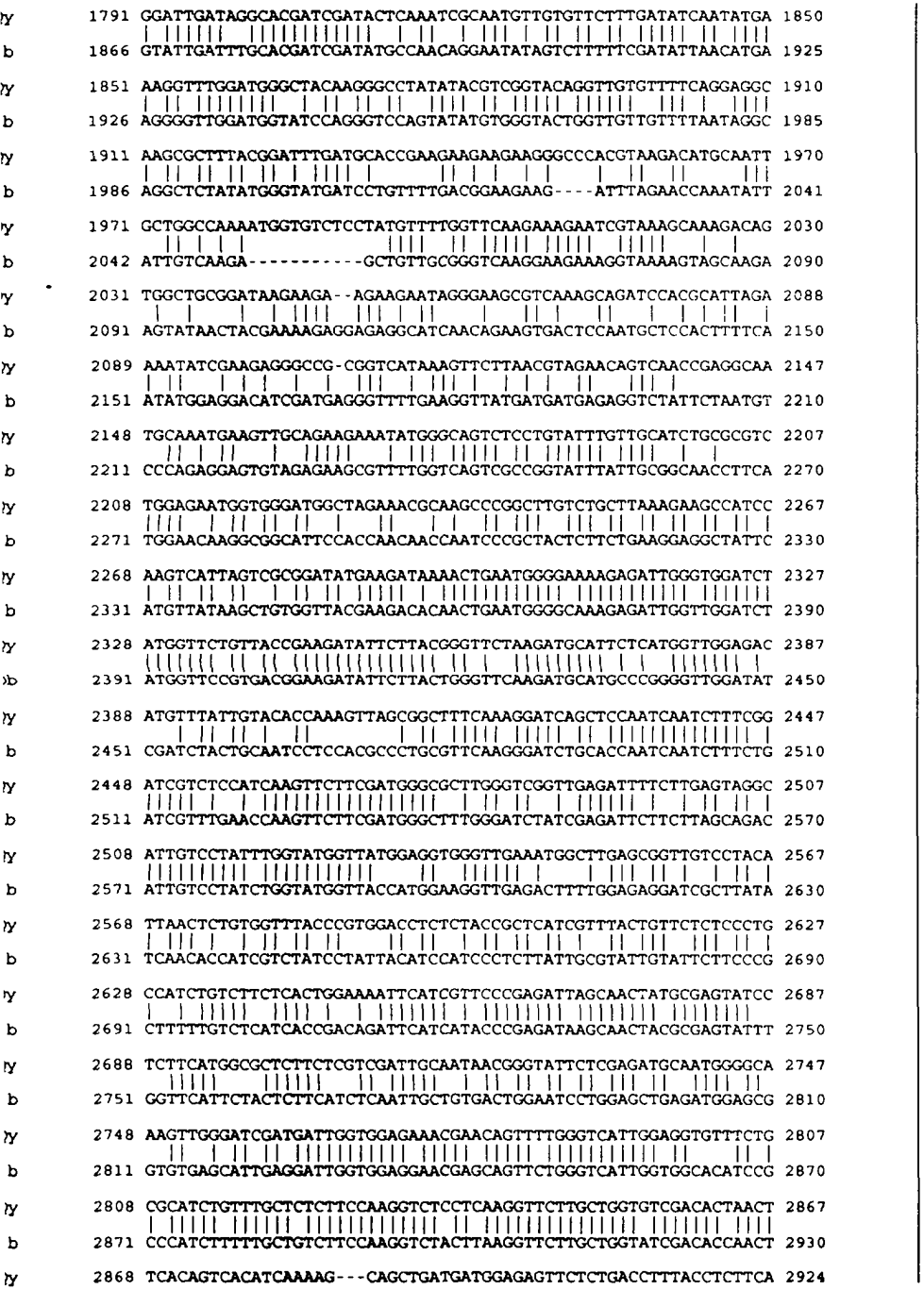
Qy       1491 GCATGATTCAGGTCTTCCTTGGAAGTGACGGTGTTCGTGATGTCGAAAACAACGAGTTGC 1550
              | ||||| ||||| ||| | ||  |  || | ||  |||  |||   || ||| |||
Db       1566 GAATGATACAGGTGTTCTTAGGCCATAGTGGGGGTCTGGATACCGATGGAAATGAGCTGC 1625

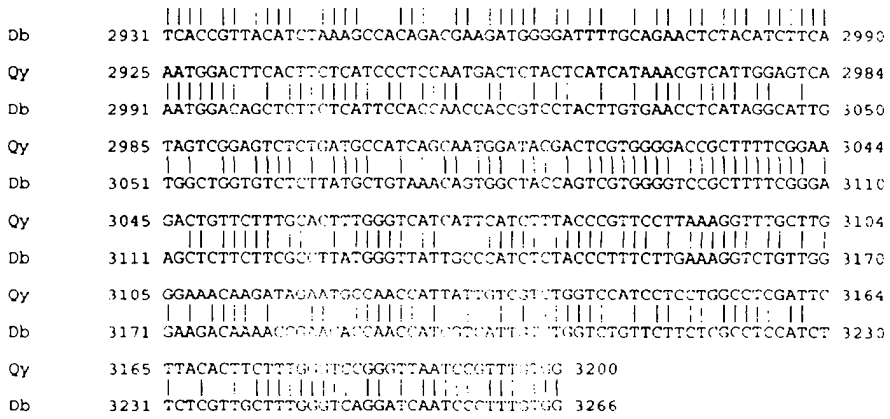
Qy       1551 CTCGATTAGTTTACGTTTCTCGTGAGAAGAGACCCGGATTTGATCACCATAAGAAGGCTG 1610
              || || |  | || |||||||||| ||| | || |||||| | ||||| || |||||||
Db       1626 CTAGACTCATCTATGTTTCTCGTGAAAAGCGGCCTGGATTTCAACACCACAAAAAGGCTG 1685

Qy       1611 GAGCTATGAATTCCCTGATACGAGTCTCTGGGGTTCTATCAAATGCTCCTTACCTTCTGA 1670
              |||||||||| |  |||| |||||  | ||||   |||| |  || ||| |||
Db       1686 GAGCTATGAATGCATTGATCCGTGTATCTGTTGTTCTTACCAATGGAGCATATCTTTTGA 1745

Qy       1671 ATGTCGATTGTGATCACTACATCAACAATAGCAAAGCTCTTAGAGAAGCAATGTGTTTCA 1730
              | || |||||||||| ||| | || || || ||| ||| |||||| ||||||||||
Db       1746 ACGTGGATTGTGATCATTACTTTAATAACAGTAAGGCTATTAAAGAAGCTATGTGTTTCA 1805

Qy       1731 TGATGGATCCTCAGTCAGGAAAGAAAATCTGTTATGTTCAGTTCCCTCAAAGGTTCGATG 1790
              |||||||  ||       ||||||||   ||| ||||| |||||||||||| | || || |
Db       1806 TGATGGACCCGGCTATTGGAAAGAAGTGCTGCTATGTCCAGTTCCCTCAACGTTTTGACG 1865
```





Search completed: October 30, 2003, 12:47:16
Job time : 848.27 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 30, 2003, 05:33:15 ; Search time 12271.2 Seconds
(without alignments)
11481.594 Million cell updates/sec

Title: US-09-980-043-5
Perfect score: 3444
Sequence: 1 gcggccgcggttaatcgccg.....gtgttgtagatagcggccgc 3444

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3444	100.0	3444	8	AF062485	AF062485 Arabidops
2	3211.2	93.2	3255	6	AX507835	AX507835 Sequence
3	2096.4	60.9	39645	8	AB025637	AB025637 Arabidops
4	1944	56.4	3255	6	AX505864	AX505864 Sequence
5	1944	56.4	3771	8	AF027173	AF027173 Arabidops
6	1934.4	56.2	3828	6	AR267558	AR267558 Sequence
7	1934.4	56.2	3828	6	AX030944	AX030944 Sequence
8	1934.4	56.2	3828	6	BD022677	BD022677 Manipulat
9	1758	51.0	3847	8	AF304374	AF304374 Nicotiana
10	1489.8	43.3	3968	8	AF200531	AF200531 Zea mays
11	1462.4	42.5	3538	8	AF200530	AF200530 Zea mays
12	1456	42.3	1492	8	AF375439	AF375439 Arabidops
13	1416.4	41.1	3812	8	AF200532	AF200532 Zea mays
14	1294.8	37.6	3603	6	AR267557	AR267557 Sequence
15	1294.8	37.6	3603	6	AX030942	AX030942 Sequence
16	1294.8	37.6	3603	6	BD022676	BD022676 Manipulat
17	1294.8	37.6	3673	6	AR267560	AR267560 Sequence
18	1294.8	37.6	3673	6	AX030948	AX030948 Sequence
19	1294.8	37.6	3673	6	BD022679	BD022679 Manipulat
20	1294.8	37.6	3763	8	BT008654	BT008654 Arabidops
21	1273.8	37.0	3723	8	AF150630	AF150630 Gossypium
22	1257.4	36.5	3229	8	BT002335	BT002335 Arabidops
23	1255.6	36.5	3682	8	AF027174	AF027174 Arabidops
24	1252.6	36.4	3614	6	AR267559	AR267559 Sequence
25	1252.6	36.4	3614	6	AX030946	AX030946 Sequence
26	1252.6	36.4	3614	6	BD022678	BD022678 Manipulat
27	1216.2	35.3	82390	8	AB016893	AB016893 Arabidops
28	1212.8	35.2	3725	8	AF200526	AF200526 Zea mays
29	1205.2	35.0	3752	8	AF200525	AF200525 Zea mays
30	1204.4	35.0	3676	8	AF200529	AF200529 Zea mays
31	1201.8	34.9	3222	6	AX652952	AX652952 Sequence
32	1178.2	34.2	3264	6	AX653232	AX653232 Sequence
33	1173.4	34.1	3395	8	AF081534	AF081534 Populus x
34	1169.8	34.0	3277	8	AY095297	AY095297 Populus t
35	1168.4	33.9	3795	8	AF200533	AF200533 Zea mays
36	1168.4	33.9	3799	6	AX338680	AX338680 Sequence
37	1160	33.7	3745	8	AF200528	AF200528 Zea mays
38	1153	33.5	3081	8	AF088917	AF088917 Arabidops
39	1151.4	33.4	3081	8	BT004543	BT004543 Arabidops
40	1151.4	33.4	3355	8	AY139754	AY139754 Arabidops
41	1145	33.2	3186	8	GHU58283	U58283 Gossypium h
42	1145	33.2	3328	6	AR164434	AR164434 Sequence
43	1137.4	33.0	2830	6	AX338676	AX338676 Sequence
44	1137	33.0	2828	8	AF200527	AF200527 Zea mays
45	1126.8	32.7	3168	8	AF458083	AF458083 Arabidops

ALIGNMENTS

RESULT 1
AF062485
LOCUS AF062485 3444 bp mRNA linear PLN 08-AUG-1998
DEFINITION Arabidopsis thaliana cellulose synthase mRNA, partial cds.
ACCESSION AF062485
VERSION AF062485.1 GI:3135610
KEYWORDS .
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 3444)
AUTHORS Wu,L., Joshi,C.P. and Chiang,V.L.
TITLE AraxCelA, a new member of the cellulose synthase gene family from

```
          Arabidopsis (Accession No. AF062485) (PGR98-114)
 JOURNAL  Plant Physiol. 117 (3), 1125 (1998)
EFERENCE  2  (bases 1 to 3444)
 AUTHORS  Joshi,C.
 TITLE    Direct Submission
 JOURNAL  Submitted (01-MAY-1998) Plant Biotechnology Research Center, School
          of Forestry and Wood Products, Michigan Technological University,
          Houghton, MI 49931, USA
EATURES             Location/Qualifiers
     source         1. .3444
                    /organism="Arabidopsis thaliana"
                    /mol_type="mRNA"
                    /cultivar="Columbia"
                    /db_xref="taxon:3702"
                    /clone_lib="ABRC Number CD4-16"
     CDS            <1. .3247
                    /codon_start=2
                    /product="cellulose synthase"
                    /protein_id="AAC29067.1"
                    /db_xref="GI:3135611"
                    /translation="RPRLIAGSHNRNEFVLINADENARIRSVQELSGQTCQICKDEIE
                    LTVDGEPFVACNECAFPVCRPCYEYERREGNQACPQCKTRFKRLKGSPRVEGDEEEDD
                    IDDLDNEFEYGNNGIGFDQVSEGMSISRRNSGFPQSDLDSAPPGSQIPLLTYGDEDVE
                    ISSDRHALIVPPSLGGHGNRVHPVSLSDPTVAAHRRLMVPQKDLAVYGYGSVAWKDRM
                    EEWKRKQNEKLQVVRHEGDPDFEDGDDADFPMMDEGRQPLSMKIPIKSSKINPYRMLI
                    VLRLVILGLFFHYRILHPVKDAYALWLISVICEIWFAVSWVLDQFPKWYPIERETYLD
                    RLSLRYEKEGKPSGLSPVDVFVSTVDPLKEPPLITANTVLSILAVDYPVDKVACYVSD
                    DGAAMLTFEALSETAEFARKWVPFCKKYCIEPRAPEWYFCHKMDYLKNKVHPAFVRER
                    RAMKRDYEEFKVKINALVATAQKVPEDGWTMQDGTPWPGNSVRDHPGMIQVFLGSDGV
                    RDVENNELPRLVYVSREKRPGFDHHKKAGAMNSLIRVSGVLSNAPYLLNVDCDHYINN
                    SKALREAMCFMMDPQSGKKICYVQFPQRFDGIDRHDRYSNRNVVFFDINMKGLDGLQG
                    PIYVGTGCVFRRQALYGFDAPKKKKGPRKTCNCWPKWCLLCFGSRKNRKAKTVAADKK
                    KKNREASKQIHALENIEEGRGHKVLNVEQSTEAMQMKLQKKYGQSPVFVASARLENGG
                    MARNASPACLLKEAIQVISRGYEDKTEWGKEIGWIYGSVTEDILTGSKMHSHGWRHVY
                    CTPKLAAFKGSAPINLSDRLHQVLRWALGSVEIFLSRHCPIWYGYGGGLKWLERLSYI
                    NSVVYPWTSLPLIVYCSLPAICLLTGKFIVPEISNYASILFMALFSSIAITGILEMQW
                    GKVGIDDWWRNEQFWVIGGVSAHLFALFQGLLKVLAGVDTNFTVTSKAADDGEFSDLY
                    LFKWTSLLIPPMTLLIINVIGVIVGVSDAISNGYDSWGPLFGRLFFALWVIIHLYPFL
                    KGLLGKQDRMPTIIVVWSILLASILTLLWVRVNPFVAKGGPILEICGLDCL"
ASE COUNT      903 a    664 c    860 g   1017 t
RIGIN

 Query Match           100.0%;  Score 3444;  DB 8;  Length 3444;
 Best Local Similarity  100.0%;  Pred. No. 0;
 Matches 3444;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GCGGCCGCGGTTAATCGCCGGTTCTCACAACAGGAATGAGTTTGTCCTCATTAATGCCGA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GCGGCCGCGGTTAATCGCCGGTTCTCACAACAGGAATGAGTTTGTCCTCATTAATGCCGA 60

Qy         61 TGAGAATGCCCGAATAAGATCAGTCCAAGAGCTGAGTGGACAGACATGTCAAATCTGCAG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TGAGAATGCCCGAATAAGATCAGTCCAAGAGCTGAGTGGACAGACATGTCAAATCTGCAG 120

Qy        121 AGATGAGATCGAATTGACTGTTGATGGAGAACCGTTTGTGGCATGTAACGAATGTGCATT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AGATGAGATCGAATTGACTGTTGATGGAGAACCGTTTGTGGCATGTAACGAATGTGCATT 180

Qy        181 CCCTGTGTGTAGACCTTGCTATGAGTACGAAAGACGAGAAGGCAATCAAGCTTGTCCACA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCCTGTGTGTAGACCTTGCTATGAGTACGAAAGACGAGAAGGCAATCAAGCTTGTCCACA 240

Qy        241 GTGCAAAACCCGTTTCAAACGTCTTAAAGGAAGTCCAAGAGTTGAAGGTGATGAAGAGGA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTGCAAAACCCGTTTCAAACGTCTTAAAGGAAGTCCAAGAGTTGAAGGTGATGAAGAGGA 300

Qy        301 AGATGACATTGATGATTTAGACAATGAGTTTGAGTATGGAAATAATGGGATTGGATTTGA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGATGACATTGATGATTTAGACAATGAGTTTGAGTATGGAAATAATGGGATTGGATTTGA 360

Qy        361 TCAGGTTTCTGAAGGTATGTCAATCTCTCGTCGCAACTCCGGTTTCCCACAATCTGATTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TCAGGTTTCTGAAGGTATGTCAATCTCTCGTCGCAACTCCGGTTTCCCACAATCTGATTT 420

Qy        421 GGATTCAGCTCCAC TGGCTCTCAGATTCCATTGCTGACTTACGGCGACGAGGACGTTGA 480
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGATTCAGCTCCAC TGGCTCTCAGATTCCATTGCTGACTTACGGCGACGAGGACGTTGA 480

Qy        481 GATTTCTTCTGATA ACATGCTCTTATTGTTCCTCCTTCACTTGGTGGTCATGGCAATAG 540
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        481 GATTTCTTCTGATA ACATGCTCTTATTGTTCCTCCTTCACTTGGTGGTCATGGCAATAG 540

Qy        541 AGTTCATCCTGTTT TCTTTCTGACCCGACCGTGGCTGCACATCGAAGGCTGATGGTACC 600
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        541 AGTTCATCCTGTTT TCTTTCTGACCCGACCGTGGCTGCACATCGAAGGCTGATGGTACC 600

Qy        601 TCAGAAAGATCTT GGTTTATGGTTATGGAAGTGTCGCTTGGAAAGATCGGATGGAGGA 660
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        601 TCAGAAAGATCTT GGTTTATGGTTATGGAAGTGTCGCTTGGAAAGATCGGATGGAGGA 660

Qy        661 ATGGAAGAGAAAA AAATGAGAAACTTAGCTTGTTAGGCATGAAGGAGATCCTGATTT 720
              ||| ||| ||| ||| ||| | | || ||||| |||| |||| |||| |||| | |
Db        661 ATGGAAGAGAAAA AAATGAGAAACTTAGCTTGTTAGGCATGAAGGAGATCCTGATTT 720

Qy        721 TGAAGATGGTGATGATGCTGATTTTCCAATGATGGATGAGGGAAGGCAGCCATTGTCTAT 780
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TGAAGATGGTGATGATGCTGATTTTCCAATGATGGATGAGGGAAGGCAGCCATTGTCTAT 780

Qy        781 GAAGATACCAATCAAATCGAGCAAGATAAATCGTTACCGGATGTTAATTGTGCTACGTCT 840
              |||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAAGATACCAATCAAATCGAGCAAGATAAATCGTTACCGGATGTTAATTGTGCTACGTCT 840

Qy        841 TGTGATTCTTGGTCTCTTCTTTCACTACCGTATTCTTCACCCCGTCAAAGATGCATATGC 900
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TGTGATTCTTGGTCTCTTCTTTCACTACCGTATTCTTCACCCCGTCAAAGATGCATATGC 900

Qy        901 TTTGTGGCTTATTT TGTTATATGTGAGATATGGTTTGCTGTTTCATGGGTTCTTGATCA 960
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTTGTGGCTTATTT TGTTATATGTGAGATATGGTTTGCTGTTTCATGGGTTCTTGATCA 960

Qy        961 GTTCCCTAAATGGT ACCCTATCGAGCGAGAAACGTACTTGGACCGACTCTCATTAAGATA 1020
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        961 GTTCCCTAAATGGTACCCTATCGAGCGAGAAACGTACTTGGACCGACTCTCATTAAGATA 1020

Qy       1021 TGAGAAAGAAGGGAAACCGTCGGGACTATCCCCTGTGGATGTATTTGTTAGTACAGTGGA 1080
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 TGAGAAAGAAGGGAAACCGTCGGGACTATCCCCTGTGGATGTATTTGTTAGTACAGTGGA 1080

Qy       1081 TCCATTGAAAGAGCGTCCGCTTATTACTGCAAATACTGTCTTGTCTATTCTTGCTGTTGA 1140
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 TCCATTGAAAGAGCGTCCGCTTATTACTGCAAATACTGTCTTGTCTATTCTTGCTGTTGA 1140

Qy       1141 TTATCCTGTCGATAAGGTTGCTTGTTACGTATCTGATGATGGTGCTGCTATGCTTACTTT 1200
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 TTATCCTGTCGATAAGGTTGCTTGTTACGTATCTGATGATGGTGCTGCTATGCTTACTTT 1200

Qy       1201 CGAAGCTCTTTCTGAGACCGCTGAATTCGCAAGGAAATGGGTTCCTTTCTGCAAGAAATA 1260
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 CGAAGCTCTTTCTGAGACCGCTGAATTCGCAAGGAAATGGGTTCCTTTCTGCAAGAAATA 1260

Qy       1261 TTGTATTGAGCCTCGTGCTCCCGAATGGTATTTCTGCCATAAAATGGACTACTTGAAGAA 1320
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 TTGTATTGAGCCTCGTGCTCCCGAATGGTATTTCTGCCATAAAATGGACTACTTGAAGAA 1320

Qy       1321 TAAAGTTCATCCCGCATTTGTTAGGGAGCGGCGAGCCATGAAGAGAGATTATGAAGAATT 1380
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TAAAGTTCATCCCGCATTTGTTAGGGAGCGGCGAGCCATGAAGAGAGATTATGAAGAATT 1380

Qy       1381 CAAAGTAAAGATCAATGCTTTAGTAGCAACAGCACAGAAAGTGCCTGAGGATGGTTGGAC 1440
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CAAAGTAAAGATCAATGCTTTAGTAGCAACAGCACAGAAAGTGCCTGAGGATGGTTGGAC 1440

Qy       1441 TATGCAAGACGGTACACCTTGGCCCGGTAATAGTGTGCGAGATCATCCTGGCATGATTCA 1500
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 TATGCAAGACGGTACACCTTGGCCCGGTAATAGTGTGCGAGATCATCCTGGCATGATTCA 1500
```

```
Qy       1501 GGTCTTCCTTGGAAGTGACGGTGTTCGTGATGTCGAAAACAACGAGTTGCCTCGATTAGT 1560
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 GGTCTTCCTTGGAAGTGACGGTGTTCGTGATGTCGAAAACAACGAGTTGCCTCGATTAGT 1560

Qy       1561 TTACGTTTCTCGTGAGAAGAGACCCGGATTTGATCACCATAAGAAGGCTGGAGCTATGAA 1620
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 TTACGTTTCTCGTGAGAAGAGACCCGGATTTGATCACCATAAGAAGGCTGGAGCTATGAA 1620

Qy       1621 TTCCCTGATACGAGTCTCTGGGGTTCTATCAAATGCTCCTTACCTTCTGAATGTCGATTG 1680
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TTCCCTGATACGAGTCTCTGGGGTTCTATCAAATGCTCCTTACCTTCTGAATGTCGATTG 1680

Qy       1681 TGATCACTACATCAACAATAGCAAAGCTCTTAGAGAAGCAATGTGTTTCATGATGGATCC 1740
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TGATCACTACATCAACAATAGCAAAGCTCTTAGAGAAGCAATGTGTTTCATGATGGATCC 1740

Qy       1741 TCAGTCAGGAAAGAAAATCTGTTATGTTCAGTTCCCTCAAAGGTTCGATGGGATTGATAG 1800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 TCAGTCAGGAAAGAAAATCTGTTATGTTCAGTTCCCTCAAAGGTTCGATGGGATTGATAG 1800

Qy       1801 GCACGATCGATACTCAAATCGCAATGTTGTGTTCTTTGATATCAATATGAAAGGTTTGGA 1860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 GCACGATCGATACTCAAATCGCAATGTTGTGTTCTTTGATATCAATATGAAAGGTTTGGA 1860

Qy       1861 TGGGCTACAAGGGCCTATATACGTCGGTACAGGTTGTGTTTTCAGGAGGCAAGCGCTTTA 1920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 TGGGCTACAAGGGCCTATATACGTCGGTACAGGTTGTGTTTTCAGGAGGCAAGCGCTTTA 1920

Qy       1921 CGGATTTGATGCACCGAAGAAGAAGAAGGGCCCACGTAAGACATGCAATTGCTGGCCAAA 1980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 CGGATTTGATGCACCGAAGAAGAAGAAGGGCCCACGTAAGACATGCAATTGCTGGCCAAA 1980

Qy       1981 ATGGTGTCTCCTATGTTTTGGTTCAAGAAAGAATCGTAAAGCAAAGACAGTGGCTGCGGA 2040
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 ATGGTGTCTCCTATGTTTTGGTTCAAGAAAGAATCGTAAAGCAAAGACAGTGGCTGCGGA 2040

Qy       2041 TAAGAAGAAGAAGAATAGGGAAGCGTCAAAGCAGATCCACGCATTAGAAAATATCGAAGA 2100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 TAAGAAGAAGAAGAATAGGGAAGCGTCAAAGCAGATCCACGCATTAGAAAATATCGAAGA 2100

Qy       2101 GGGCCGCGGTCATAAAGTTCTTAACGTAGAACAGTCAACCGAGGCAATGCAAATGAAGTT 2160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 GGGCCGCGGTCATAAAGTTCTTAACGTAGAACAGTCAACCGAGGCAATGCAAATGAAGTT 2160

Qy       2161 GCAGAAGAAATATGGGCAGTCTCCTGTATTTGTTGCATCTGCGCGTCTGGAGAATGGTGG 2220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 GCAGAAGAAATATGGGCAGTCTCCTGTATTTGTTGCATCTGCGCGTCTGGAGAATGGTGG 2220

Qy       2221 GATGGCTAGAAACGCAAGCCCGGCTTGTCTGCTTAAAGAAACATCCAAGTCATTAGTCG 2280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 GATGGCTAGAAACGCAAGCCCGGCTTGTCTGCTTAAAGAAACATCCAAGTCATTAGTCG 2280

Qy       2281 CGGATATGAAGATAAAACTGAATGGGGAAAAGAGATTGGGTGGATCTATGGTTCTGTTAC 2340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 CGGATATGAAGATAAAACTGAATGGGGAAAAGAGATTGGGTGGATCTATGGTTCTGTTAC 2340

Qy       2341 CGAAGATATTCTTACGGGTTCTAAGATGCATTCTCATGGTTGGAGACATGTTTATTGTAC 2400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 CGAAGATATTCTTACGGGTTCTAAGATGCATTCTCATGGTTGGAGACATGTTTATTGTAC 2400

Qy       2401 ACCAAAGTTAGCGGCTTTCAAAGGATCAGCTCCAATCAATCTTTCGGATCGTCTCCATCA 2460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 ACCAAAGTTAGCGGCTTTCAAAGGATCAGCTCCAATCAATCTTTCGGATCGTCTCCATCA 2460

Qy       2461 AGTTCTTCGATGGGCGCTTGGGTCGGTTGAGATTTTCTTGAGTAGGCATTGTCCTATTTG 2520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2461 AGTTCTTCGATGGGCGCTTGGGTCGGTTGAGATTTTCTTGAGTAGGCATTGTCCTATTTG 2520

Qy       2521 GTATGGTTATGGAGGTGGGTTGAAATGGCTTGAGCGGTTGTCCTACATTAACTCTGTGGT 2580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2521 GTATGGTTATGGAGGTGGGTTGAAATGGCTTGAGCGGTTGTCCTACATTAACTCTGTGGT 2580

Qy       2581 TTACCCGTGGACCTCTCTACCGCTCATCGTTTACTGTTCTCTCCCTGCCATCTGTCTTCT 2640
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2581 TTACCCGTGGACCTCTCTACCGCTCATCGTTTACTGTTCTCTCCCTGCCATCTGTCTTCT 2640

Qy       2641 CACTGGAAAATTCATCGTTCCCGAGATTAGCAACTATGCGAGTATCCTCTTCATGGCGCT 2700
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2641 CACTGGAAAATTCATCGTTCCCGAGATTAGCAACTATGCGAGTATCCTCTTCATGGCGCT 2700

Qy       2701 CTTCTCGTCGATTGCAATAACGGGTATTCTCGAGATGCAATGGGGCAAAGTTGGGATCGA 2760
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2701 CTTCTCGTCGATTGCAATAACGGGTATTCTCGAGATGCAATGGGGCAAAGTTGGGATCGA 2760

Qy       2761 TGATTGGTGGAGAAACGAACAGTTTTGGGTCATTGGAGGTGTTTCTGCGCATCTGTTTGC 2820
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2761 TGATTGGTGGAGAAACGAACAGTTTTGGGTCATTGGAGGTGTTTCTGCGCATCTGTTTGC 2820

Qy       2821 TCTCTTCCAAGGTCTCCTCAAGGTTCTTGCTGGTGTCGACACTAACTTCACAGTCACATC 2880
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2821 TCTCTTCCAAGGTCTCCTCAAGGTTCTTGCTGGTGTCGACACTAACTTCACAGTCACATC 2880

Qy       2881 AAAAGCAGCTGATGATGGAGAGTTCTCTGACCTTTACCTCTTCAAATGGACTTCACTTCT 2940
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2881 AAAAGCAGCTGATGATGGAGAGTTCTCTGACCTTTACCTCTTCAAATGGACTTCACTTCT 2940

Qy       2941 CATCCCTCCAATGACTCTACTCATCATAAACGTCATTGGAGTCATAGTCGGAGTCTCTGA 3000
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2941 CATCCCTCCAATGACTCTACTCATCATAAACGTCATTGGAGTCATAGTCGGAGTCTCTGA 3000

Qy       3001 TGCCATCAGCAATGGATACGACTCGTGGGGACCGCTTTTCGGAAGACTGTTCTTTGCACT 3060
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3001 TGCCATCAGCAATGGATACGACTCGTGGGGACCGCTTTTCGGAAGACTGTTCTTTGCACT 3060

Qy       3061 TTGGGTCATCATTCATCTTTACCCGTTCCTTAAAGGTTTGCTTGGGAAACAAGATAGAAT 3120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3061 TTGGGTCATCATTCATCTTTACCCGTTCCTTAAAGGTTTGCTTGGGAAACAAGATAGAAT 3120

Qy       3121 GCCAACCATTATTGTCGTCTGGTCCATCCTCCTGGCCTCGATTCTTACACTTCTTTGGGT 3180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3121 GCCAACCATTATTGTCGTCTGGTCCATCCTCCTGGCCTCGATTCTTACACTTCTTTGGGT 3180

Qy       3181 CCGGGTTAATCCGTTTGTGGCGAAAGGCGGTCCTATTCTCGAGATCTGTGGTTTAGACTG 3240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3181 CCGGGTTAATCCGTTTGTGGCGAAAGGCGGTCCTATTCTCGAGATCTGTGGTTTAGACTG 3240

Qy       3241 CTTGTGATTCGATTGACCGGTGGATGGGTTGGTGAAAAAGGTTTAATTCCCACGGATCAA 3300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3241 CTTGTGATTCGATTGACCGGTGGATGGGTTGGTGAAAAAGGTTTAATTCCCACGGATCAA 3300

Qy       3301 AGAGAGGTAAGAGAGATATTGTTTTACCTCTAAAAGACTCCTTCATTGTGTTCATTAGAT 3360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3301 AGAGAGGTAAGAGAGATATTGTTTTACCTCTAAAAGACTCCTTCATTGTGTTCATTAGAT 3360

Qy       3361 GATGAAAAATGAAAAGAAAAAAGAAGATTTAATTTTGTTACGAGAATTGTTATTTTTGCAA 3420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3361 GATGAAAAATGAAAAGAAAAAAGAAGATTTAATTTTGTTACGAGAATTGTTATTTTTGCAA 3420

Qy       3421 GAATGTGTTGTAGATAGCGGCCGC 3444
             ||||||||||||||||||||||||
Db       3421 GAATGTGTTGTAGATAGCGGCCGC 3444
```

RESULT 2
AX507835

```
LOCUS       AX507835               3255 bp    DNA     linear   PAT 27-SEP-2002
DEFINITION  Sequence 2530 from Patent WO0216655.
ACCESSION   AX507835
VERSION     AX507835.1  GI:23389072
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
```

```
 AUTHORS   Harper,J.F., Kreps,J., Wang,X. and Zhu,T.
 TITLE     Stress-regulated genes of plants, transgenic plants containing
           same, and methods of use
 JOURNAL   Patent: WO 0216655-A 2530 28-FEB-2002;
           The Scripps Research Institute (US) ; Syngenta Participations AG
           (CH)
'EATURES             Location/Qualifiers
     source          1. .3255
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3702"
)ASE COUNT      841 a    641 c    813 g    960 t
)RIGIN

 Query Match             93.2%;  Score 3211.2;  DB 6;  Length 3255;
 Best Local Similarity   99.4%;  Pred. No. 0;
 Matches 3222;  Conservative    0;  Mismatches   18;  Indels    0;  Gaps    0;

Qy          8 CGGTTAATCGCCGGTTCTCACAACAGGAATGAGTTTGTCCTCATTAATGCCGATGAGAAT 67
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         16 CGGTTAATCGCCGGTTCTCACAACAGGAATGAGTTTGTCCTCATTAATGCCGATGAGAAT 75

Qy         68 GCCCGAATAAGATCAGTCCAAGAGCTGAGTGGACAGACATGTCAAATCTGCAGAGATGAG 127
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         76 GCCCGAATAAGATCAGTCCAAGAGCTGAGTGGACAGACATGTCAAATCTGCAGAGATGAG 135

Qy        128 ATCGAATTGACTGTTGATGGAGAACCGTTTGTGGCATGTAACGAATGTGCATTCCCTGTG 187
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        136 ATCGAATTGACTGTTGATGGAGAACCGTTTGTGGCATGTAACGAATGTGCATTCCCTGTG 195

Qy        188 TGTAGACCTTGCTATGAGTACGAAAGACGAGAAGGCAATCAAGCTTGTCCACAGTGCAAA 247
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        196 TGTAGACCTTGCTATGAGTACGAAAGACGAGAAGGCAATCAAGCTTGTCCACAGTGCAAA 255

Qy        248 ACCCGTTTCAAACGTCTTAAAGGAAGTCCAAGAGTTGAAGGTGATGAAGAGGAAGATGAC 307
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        256 ACCCGTTTCAAACGTCTTAAAGGAAGTCCAAGAGTTGAAGGTGATGAAGAGGAAGATGAC 315

Qy        308 ATTGATGATTTAGACAATGAGTTTGAGTATGGAAATAATGGGATTGGATTTGATCAGGTT 367
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        316 ATTGATGATTTAGACAATGAGTTTGAGTATGGAAATAATGGGATTGGATTTGATCAGGTT 375

Qy        368 TCTGAAGGTATGTCAATCTCTCGTCGCAACTCCGGTTTCCCACAATCTGATTTGGATTCA 427
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        376 TCTGAAGGTATGTCAATCTCTCGTCGCAACTCCGGTTTCCCACAATCTGATTTGGATTCA 435

Qy        428 GCTCCACCTGGCTCTCAGATTCCATTGCTGACTTACGGCGACGAGGACGTTGAGATTTCT 487
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        436 GCTCCACCTGGCTCTCAGATTCCATTGCTGACTTACGGCGACGAGGACGTTGAGATTTCT 495

Qy        488 TCTGATAGACATGCTCTTATTGTTCCTCCTTCACTTGGTGGTCATGGCAATAGAGTTCAT 547
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        496 TCTGATAGACATGCTCTTATTGTTCCTCCTTCACTTGGTGGTCATGGCAATAGAGTTCAT 555

Qy        548 CCTGTTTCTCTTTCTGACCCGACCGTGGCTGCACATCGAAGGCTGATGGTACCTCAGAAA 607
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        556 CCTGTTTCTCTTTCTGACCCGACCGTGGCTGCACATCCAAGGCCTATGGTACCTCAGAAA 615

Qy        608 GATCTTGCGGTTTATGGTTATGGAAGTGTCGCTTGGAAAGATCGGATGGAGGAATGGAAG 667
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        616 GATCTTGCGGTTTATGGTTATGGAAGTGTCGCTTGGAAAGATCGGATGGAGGAATGGAAG 675

Qy        668 AGAAAGCAGAATGAGAAACTTCAGGTTGTTAGGCATGAAGGAGATCCTGATTTTGAAGAT 727
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        676 AGAAAGCAGAATGAGAAACTTCAGGTTGTTAGGCATGAAGGAGATCCTGATTTTGAAGAT 735

Qy        728 GGTGATGATGCTGATTTTCCAATGATGGATGAGGGAAGGCAGCCATTGTCTATGAAGATA 787
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        736 GGTGATGATGCTGATTTTCCAATGATGGATGAGGGAAGGCAGCCATTGTCTAGGAAGATA 795

Qy        788 CCAATCAAATCGAGCAAGATAAATCCTTACCGGATGTTAATTGTGCTACGTCTTGTGATT 847
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        796 CCAATCAAATCGAGCAAGATAAATCCTTACCGGATGTTAATTGTGCTACGTCTTGTGATT 855

Qy        848 CTTGGTCTCTTCTTTCACTACCGTATTCTTCACCCCGTCAAAGATGCATATGCTTTGTGG 907
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        856 CTTGGTCTCTTCTTTCACTACCGTATTCTTCACCCCGTCAAAGATGCATATGCTTTGTGG 915

Qy        908 CTTATTTCTGTTATATGTGAGATATGGTTTGCTGTTTCATGGGTTCTTGATCAGTTCCCT 967
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        916 CTTATTTCTGTTATATGTGAGATATGGTTTGCTGTTTCATGGGTTCTTGATCAGTTCCCT 975

Qy        968 AAATGGTACCCTATTGAGCGAGAAACGTACTTGGACCGACTCTCATTAAGATATGAGAAA 1027
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        976 AAATGGTACCCTATTGAGCGAGAAACGTACTTGGACCGACTCTCATTAAGATATGAGAAA 1035

Qy       1028 GAAGGGAAACCGTCTGGACTATCCCCTGTGGATGTATTTGTTAGTACAGTGGATCCATTG 1087
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1036 GAAGGGAAACCGTCTGGACTATCCCCTGTGGATGTATTTGTTAGTACAGTGGATCCATTG 1095

Qy       1088 AAAGAGCCTCCGCTATTACTGCAAATACTGTCTTGTCTATTCTTGCTGTTGATTATCCT 1147
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1096 AAAGAGCCTCCGCTATTACTGCAAATACTGTCTTGTCTATTCTTGCTGTTGATTATCCT 1155

Qy       1148 GTCGATAAGGTTGCTTGTTACGTATCTGATGATGGTGCTGCTATGCTTACTTTCGAAGCT 1207
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1156 GTCGATAAGGTTGCTTGTTACGTATCTGATGATGGTGCTGCTATGCTTACTTTCGAAGCT 1215

Qy       1208 CTTTCTGAGACCGCTGAATTCGCAAGGAAATGGGTTCCTTTCTGCAAGAAATATTGTATT 1267
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1216 CTTTCTGAGACCGCTGAATTCGCAAGGAAATGGGTTCCTTTCTGCAAGAAATATTGTATT 1275

Qy       1268 GAGCCTCGTGCTCCGGAATGGTATTTCTGCCATAAAATGGACTACTTGAAGAATAAAGTT 1327
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1276 GAGCCTCGTGCTCCGGAATGGTATTTCTGCCATAAAATGGACTACTTGAAGAATAAAGTT 1335

Qy       1328 CATCCCGCATTTGTTAGGGAGCGGCGAGCCATGAAGAGAGATTATGAAGAATTCAAAGTA 1387
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1336 CATCCCGCATTTGTTAGGGAGCGGCGAGCCATGAAGAGAGATTATGAAGAATTCAAAGTA 1395

Qy       1388 AAGATCAATGCTTTAGTAGCAACAGCACAGAAAGTGCCTGAGGATGGTTGGACTATGCAA 1447
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1396 AAGATCAATGCTTTAGTAGCAACAGCACAGAAAGTGCCTGAGGATGGTTGGACTATGCAA 1455

Qy       1448 GACGGTACACCTTGGCCCGGTAATAGTGTGCGAGATCATCCTGGCATGATTCAGGTCTTC 1507
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1456 GACGGTACACCTTGGCCCGGTAATAGTGTGCGAGATCATCCTGGCATGATTCAGGTCTTC 1515

Qy       1508 CTTGGAAGTGACGGTGTTCGTGATGTCGAAAACAACGAGTTGCCTCGATTAGTTTACGTT 1567
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1516 CTTGGAAGTGACGGTGTTCGTGATGTCGAAAACAACGAGTTGCCTCGATTAGTTTACGTT 1575

Qy       1568 TCTCGTGAGAAGAGACCCGGATTTGATCACCATAAGAAGGCTGGAGCTATGAATTCCCTG 1627
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1576 TCTCGTGAGAAGAGACCCGGATTTGATCACCATAAGAAGGCTGGAGCTATGAATTCCCTG 1635

Qy       1628 ATACGAGTCTCTGGGGTTCTATCAAATGCTCCTTACCTTCTGAATGTCGATTGTGATCAC 1687
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1636 ATACGAGTCTCTGGGGTTCTATCAAATGCTCCTTACCTTCTGAATGTCGATTGTGATCAC 1695

Qy       1688 TACATCAACAATAGCAAAGCTCTTAGAGAAGCAATGTGTTTCATGATGGATCCTCAGTCA 1747
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1696 TACATCAACAATAGCAAAGCTCTTAGAGAAGCAATGTGTTTCATGATGGATCCTCAGTCA 1755

Qy       1748 GGAAAGAAAATCTGTTATGTTCAGTTCCCTCAAAGGTTCGATGGGATTGATAGGCACGAT 1807
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1756 GGAAAGAAAATCTGTTATGTTCAGTTCCCTCAAAGGTTCGATGGGATTGATAGGCACGAT 1815

Qy       1808 CGATACTCAAATCGCAATGTTGTGTTCTTTGATATCAATATGAAAGGTTTGGATGGGCTA 1867
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1816 CGATACTCAAATCGCAATGTTGTGTTCTTTGATATCAATATGAAAGGTTTGGATGGGCTA 1875

Qy       1868 CAAGGGCCTATATACGTCGGTACAGGTTGTGTTTTCAGGAGGCAAGCGCTTTACGGATTT 1927
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1876 CAAGGGCCTATATACGTCGGTACAGGTTGTGTTTTCAGGAGGCAAGCGCTTTACGGATTT 1935
```

```
Qy       1928 GATGCACCGAAGAAGAAGAAGGGCCCACGTAAGACATGCAATTGCTGGCCAAAATGGTGT 1987
Db       1936 GATGCACCGAAGAAGAAGAAGGGCCCACGTAAGACATGCAATTGCTGGCCAAAATGGTGT 1995

Qy       1988 CTCCTATGTTTTGGTTCAAGAAAGAATCGTAAAGCAAAGACAGTGGCTGCGGATAAGAAG 2047
Db       1996 CTCCTATGTTTTGGTTCAAGAAAGAATCGTAAAGCAAAGACAGTGGCTGCGGATAAGAAG 2055

Qy       2048 AAGAAGAATAGGGAAGCGTCAAAGCAGATCCACGCATTAGAAAATATCGAAGAGGGCCGC 2107
Db       2056 AAGAAGAATAGGGAAGCGTCAAAGCAGATCCACGCATTAGAAAATATCGAAGAGGGCCGC 2115

Qy       2108 GGTCATAAAGTTCTTAACGTAGAACAGTCAACCGAGGCAATGCAAATGAAGTTGCAGAAG 2167
Db       2116 GTCACTAAAGGTTCTAACGTAGAACAGTCAACCGAGGCAATGCAAATGAAGTTGGAGAAG 2175

Qy       2168 AAATATGGGCAGTCTCCTGTATTTGTTGCATCTGCGCGTCTGGAGAATGGTGGGATGGCT 2227
Db       2176 AAATTTGGGCAGTCTCCTGTATTTGTTGCATCTGCGCGTATGGAGAATGGTGGGATGGCT 2235

Qy       2228 AGAAACGCAAGCCCGGCTTGTCTGCTTAAAGAAGCCATCCAAGTCATTAGTCGCGGATAT 2287
Db       2236 AGAAACGCAAGCCCGGCTTGTCTGCTTAAAGAAGCCATCCAAGTCATTAGTTGCGGATAT 2295

Qy       2288 GAAGATAAAACTGAATGGGGAAAAGAGATTGGGTGGATCTATGGTTCTGTTACCGAAGAT 2347
Db       2296 GAAGATAAAACTGAATGGGGAAAAGAGATTGGGTGGATCTATGGTTCTGTTACCGAAGAT 2355

Qy       2348 ATTCTTACGGGTTCTAAGATGCATTCTCATGGTTGGAGACATGTTTATTGTACACCAAAG 2407
Db       2356 ATTCTTACGGGTTTTAAGATGCATTCTCATGGTTGGAGATCTGTTTATTGTACACCAAAG 2415

Qy       2408 TTAGCGGCTTTCAAAGGATCAGCTCCAATCAATCTTTCGGATCGTCTCCATCAAGTTCTT 2467
Db       2416 TTAGCGGCTTTCAAAGGATCAGCTCCAATCAATCTTTCGGATCGTCTCCATCAAGTTCTT 2475

Qy       2468 CGATGGGCGCTTGGGTCGGTTGAGATTTTCTTGAGTAGGCATTGTCCTATTTGGTATGGT 2527
Db       2476 CGATGGGCGCTTGGGTCGGTTGAGATTTTCTTGAGTAGGCATTGTCCTATTTGGTATGGT 2535

Qy       2528 TATGGAGGTGGGTTGAAATGGCTTGAGCGGTTGTCCTACATTAACTCTGTGGTTTACCCG 2587
Db       2536 TATGGAGGTGGGTTGAAATGGCTTGAGCGGTTGTCCTACATTAACTCTGTGGTTTACCCG 2595

Qy       2588 TGGACCTCTCTACCGCTCATCGTTTACTGTTCTCTCCCTGCCATCTGTCTTCTCACTGGA 2647
Db       2596 TGGACCTCTCTACCGCTCATCGTTTACTGTTCTCTCCCTGCCATCTGTCTTCTCACTGGA 2655

Qy       2648 AAATTCATCGTTCCCGAGATTAGCAACTATGCGAGTATCCTCTTCATGGCGCTCTTCTCG 2707
Db       2656 AAATTCATCGTTCCCGAGATTAGCAACTATGCGAGTATCCTCTTCATGGCGCTCTTCTCG 2715

Qy       2708 TCGATTGCAATAACGGGTATTCTCGAGATGCAATGGGGCAAAGTTGGGATCGATGATTGG 2767
Db       2716 TCGATTGCAATAACGGGTATTCTCGAGATGCAATGGGGCAAAGTTGGGATCGATGATTGG 2775

Qy       2768 TGGAGAAACGAACAGTTTTGGGTCATTGGAGGTGTTTCTGCGCATCTGTTTGCTCTCTTC 2827
Db       2776 TGGAGAAACGAACAGTTTTGGGTCATTGGAGGTGTTTCTGCGCATCTGTTTGCTCTCTTC 2835

Qy       2828 CAAGGTCTCCTCAAGGTTCTTGCTGGTGTCGACACTAACTTCACAGTCACATCAAAAGCA 2887
Db       2836 CAAGGTCTCCTCAAGGTTCTTGCTGGTGTCGACACTAACTTCACAGTCACATCAAAAGCA 2895

Qy       2888 GCTGATGATGGAGAGTTCTCTGACCTTTACCTCTTCAAATGGACTTCACTTCTCATCCCT 2947
Db       2896 GCTGATGATGGAGAGTTCTCTGACCTTTACCTCTTCAAATGGACTTCACTTCTCATCCCT 2955

Qy       2948 CCAATGACTCTACTCATCATAAACGTCATTGGAGTCATAGTCGGAGTCTCTGATGCCATC 3007
Db       2956 CCAATGACTCTACTCATCATAAACGTCATTGGAGTCATAGTCGGAGTCTCTGATGCCATC 3015

Qy       3008 AGCAATGGATACGACTCGTGGGGACCGCTTTTCGGAAGACTGTTCTTTGCACTTTGGGTC 3067
Db       3016 AGCAATGGATACGACTCGTGGGGACCGCTTTTCGGAAGACTGTTCTTTGCACTTTGGGTC 3075

Qy       3068 ATCATTCATCTTTACCCGTTCCTTAAAGGTTTGCTTGGGAAACAAGATAGAATGCCAACC 3127
Db       3076 ATCATTCATCTTTACCCGTTCCTTAAAGGTTTGCTTGGGAAACAAGATAGAATGCCAACC 3135

Qy       3128 ATTATTGTCGTCTGGTCCATCCTCCTGGCCTCGATTCTTACACTTCTTTGGGTCCGGGTT 3187
Db       3136 ATTATTGTCGTCTGGTCCATCCTCCTGGCCTCGATTCTTACACTTCTTTGGGTCCGGGTT 3195

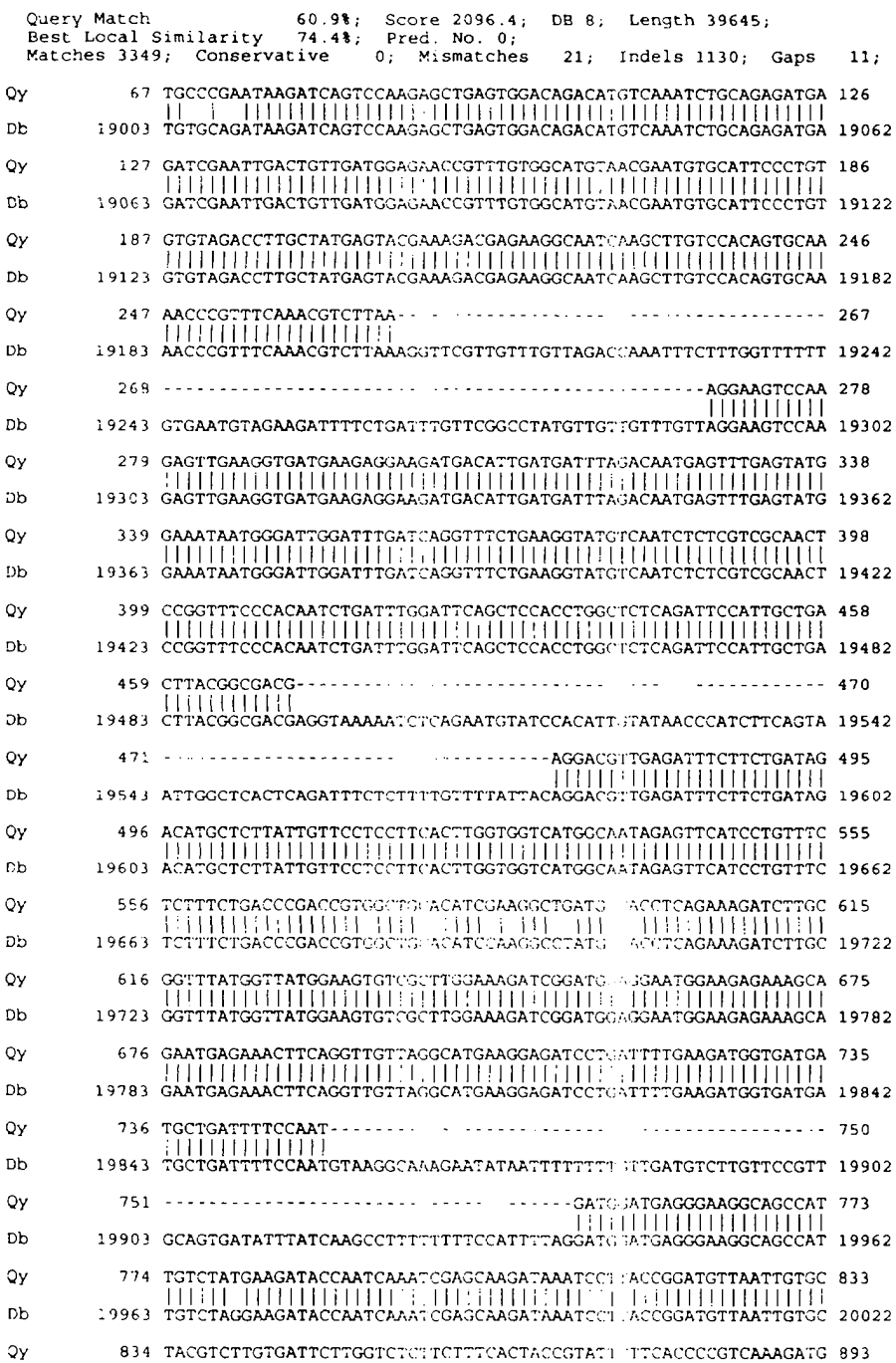
Qy       3188 AATCCGTTTGTGGCGAAAGGCGGTCCTATTCTCGAGATCTGTGGTTTAGACTGCTTGTGA 3247
Db       3196 AATCCGTTTGTGGCGAAAGGCGGTCCTATTCTCGAGATCTGTGGTTTAGACTGCTTGTGA 3255
```

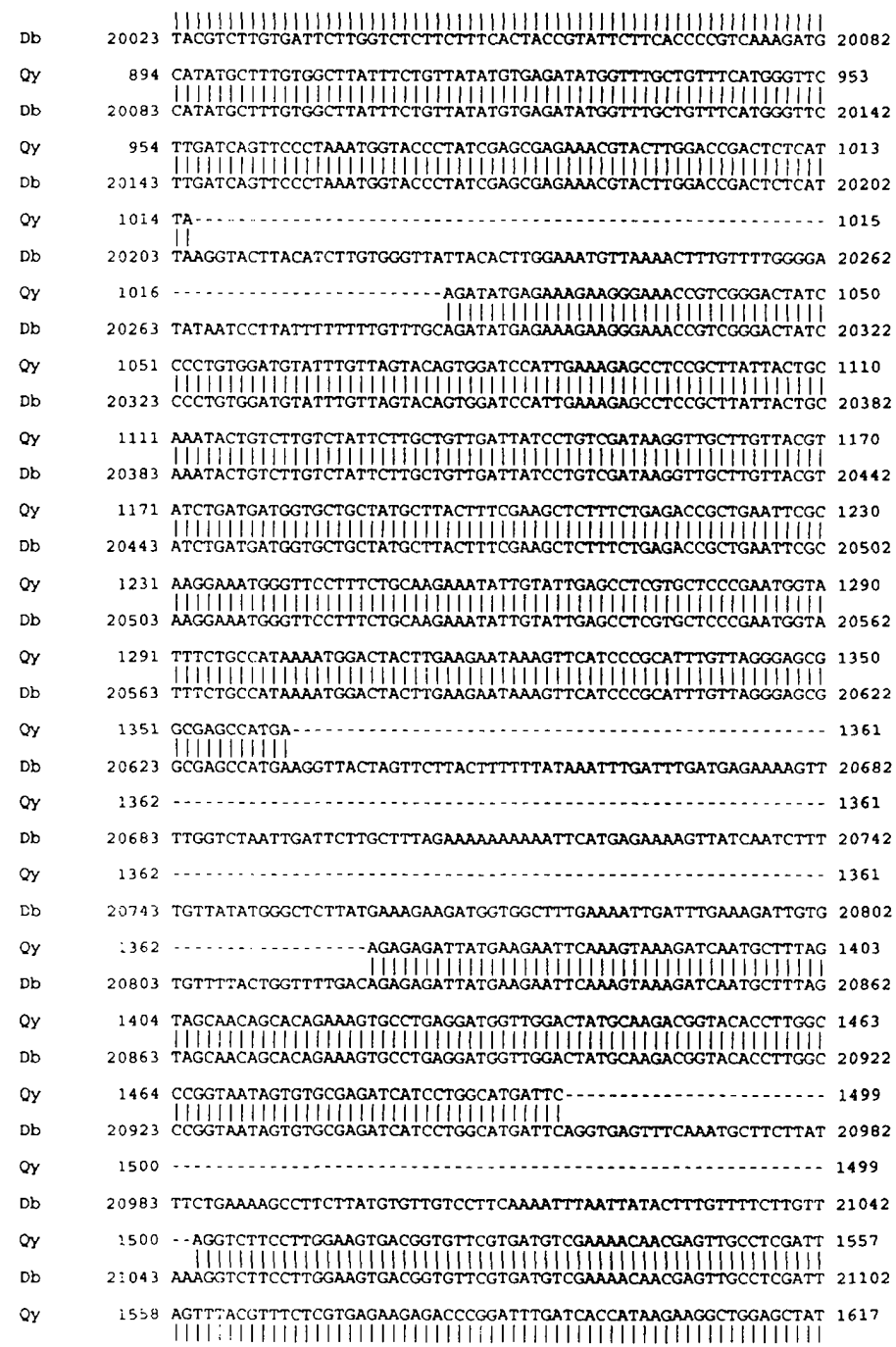
RESULT 3
AB025637
LOCUS AB025637 39645 bp DNA linear PLN 27-DEC-2000
DEFINITION Arabidopsis thaliana genomic DNA, chromosome 5, P1 clone:MVP7.
ACCESSION AB025637 BA000015
VERSION AB025637.1 GI:4589443
KEYWORDS .
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (sites)
AUTHORS Kaneko,T., Katoh,T., Asamizu,E., Sato,S., Nakamura,Y., Kotani,H. and Tabata,S.
TITLE Structural analysis of Arabidopsis thaliana chromosome 5. XI
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 39645)
AUTHORS Nakamura,Y.
TITLE Direct Submission
JOURNAL Submitted (02-APR-1999) Yasukazu Nakamura, Kazusa DNA Research Institute, Department of Plant Gene Research; 1532-3, Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:ynakamu@kazusa.or.jp, Tel:81-438-52-3935, Fax:81-438-52-3934)
COMMENT Address for correspondence: kaos@kazusa.or.jp
For the latest information on annotation of this clone, please see http://www.kazusa.or.jp/kaos/cgi-bin/agd_graph.cgi?c=MVP7
Genes with similarity to proteins in the databases are described in 'product' or 'note' qualifiers. Genes that have no significant protein similarity are described as 'unknown protein'.
The software programs used to predict genes include: Grail (Informatics Group, Oak Ridge National Laboratory, http://compbio.ornl.gov/Grail-1.3/), GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html), NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and SplicePredictor (Volker Brendel, Stanford University, http://gremlin1.zool.iastate.edu/cgi-bin/sp.cgi).
Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, Washington University School of Medicine, St. Louis, http://genome.wustl.edu/eddy/tRNAscan-SE/).
This sequence may not be the entire insert of this clone. It may be shorter because we remove overlaps between neighboring submissions. The 5' clone is MUB3 and the 3' clone is MXK3.

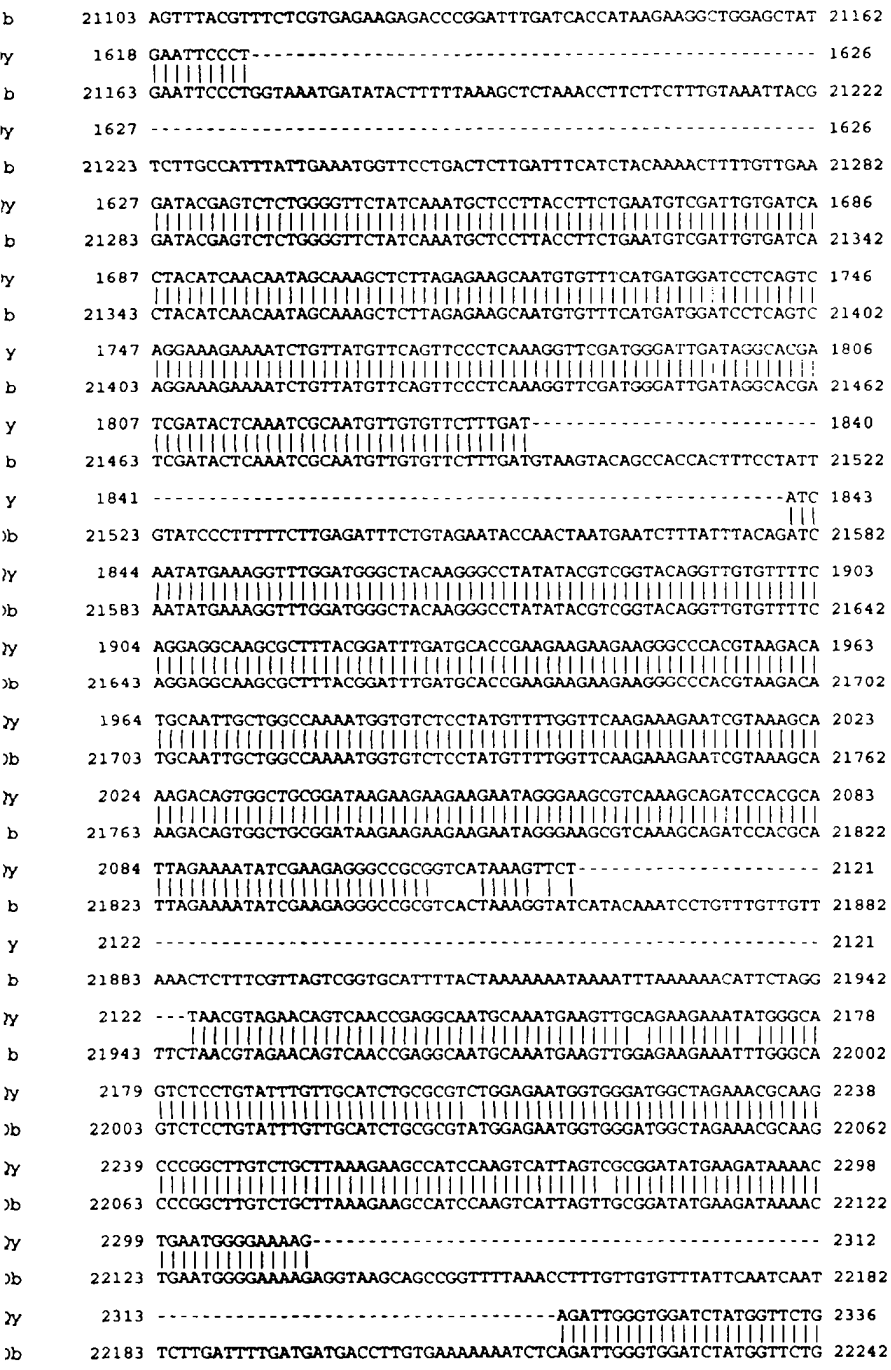
```
FEATURES             Location/Qualifiers
     source          1. .39645
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="5"
                     /clone="MVP7"
                     /clone_lib="Mitsui P1"
     CDS             complement(join(2170. .2317,2399. .2550,2646. .2804,
                     2919. .3180,3283. .3399,3505. .3567,3647. .3825))
                     /note="gene_id:MVP7.2"
                     /codon_start=1
```

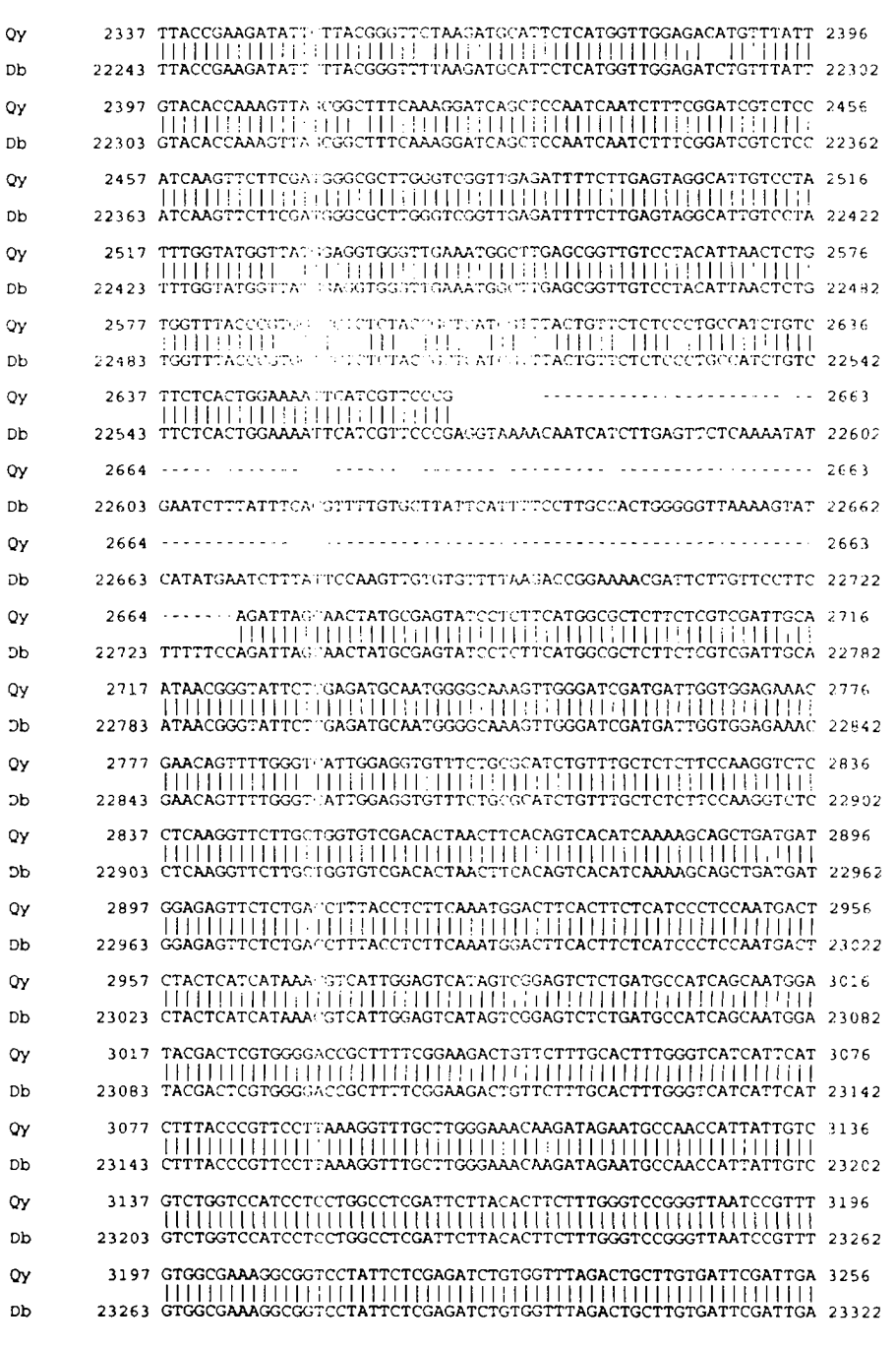
```
                /evidence=not_experimental
                /product="nodulin-like protein"
                /protein_id="BAB10303.1"
                /db_xref="GI:10177201"
                /translation="MDMESKKPYLMVTIIQVIYTIMFLISKAVFNGGMNTFVFVFYRQ
                AFATIFLAPLAFFPERKSAPPLSFVTFIKIFMLSLFGVTLSLDLNGIALSYTSATLAA
                ATTASLPAITFFLALLFGMERLKVKSIQGTAKLVGITVCMGGVIILAIYKGPLLKLPL
                CPHFYHGQEHPHRNNPGHVSGGSTSWLKGCVLMITSNILWGLWLVLQGRVLKVYPSKL
                YPTTLHCLLSSIQSFVIAIALERDISAWKLGWNLRLVAVIYCGFIVTGVAYYLQSWVI
                EKRGPVFLSMFTPLSLLFTLLSSAILLCEIISLGSIVGGLLLIIGLYCVLWGKSREEK
                NSGDDKIDLQKENDVVCNEVKVVIS"
CDS             join(6348. .6738,6847. .8146,8238. .9072)
                /note="emb|CAB71880.1
                gene_id:MVP7.3
                similar to unknown protein"
                /codon_start=1
                /evidence=not_experimental
                /protein_id="BAB10304.1"
                /db_xref="GI:10177202"
                /translation="MQFGSSKTLFFRVAGFSSTSVFKPRRAISVPVGNFSSSSSISAH
                YTSLKRSQYEEDSRSVRVPVWWDFENCHLPSGANVFKLAQTITSAVRICGIKGPITIT
                AYGDMIQLSRTNQEALFATGINLTHVPQGGKNSTDRSLITEIMCWVSQNPPPAHLFLI
                SSDSDFANVLHRLRMRNYNILLACYEETTLGVLCSAASIMWDWDALVRGQNPTAKHFN
                QPPDGPYHSWYGHYTTPLLDPFATSTNNKQISSTSVKTVELLELGSSKSRRPIPNKVV
                KQIGLILRWYPKGAAITELREQLRKRKVLLDRDFYGYKSFSRFLLSMRNILQVVPLGD
                GMFSIHAVTQGMDNKALLPKVSCENHAVVSVEKMCQNMKQNDKDVKEESHQLQESSQE
                FVQVMKLMDVKAKEEPVKANQLAITAVDDVSSFEEKDGFLKKLNRLLFGSPEMELEHL
                QERKHISGNGVLGEGKVVDKDLESQIASSTSSESAEEVKVDNAVGNGKSKSPGLTCRL
                LKRFKFSWGRYTELSNAAATGPQVDDVFVKDSFWNDVESFINSPRGFVAVSHSRSRET
                MAKNLKEEGPSSLKPLDVPKMLDLISMLISEKKWIQENPSDALPFRVTRFTEESSCRS
                NPRTTDGLRAIFVNMSESLCDGANGDKNSTNVGMSQKPKERSRSKVIADCHKLIKKIT
                EENAGGYSITKFKKAFLEKFGYRLEYRKFGFSKLQSLIEMMPEARIESGHIVTSSTPV
                PCESDSSFEDLGPVSKKIHENESSVSEGEDYDSEMEEKASSKQSGGERKKKEDETESD
                LLQILGSWDTDKKPAKTFGEDKLVEGILMSLRKKPSGDSRIQD"
CDS             complement(9326. .9793)
                /note="gb|AAD20394.1
                gene_id:MVP7.4
                similar to unknown protein"
                /codon_start=1
                /evidence=not_experimental
                /protein_id="BAB10305.1"
                /db_xref="GI:10177203"
                /translation="MATKSTSKPLLLSFLMMSYLISTFHVITVAEGRTLQFTKMATDH
                SGAGNLMDCWNAGLELKSCTDEIVKFFLSQTGTSEPPVKGGIDKDCCGAIGLVVKDCW
                SVMFTSLGLTTMEGNNLREYCEPQAEKSELSPSPAPETLALSPVEITYPGLDY"
CDS             join(10126. .10334,10710. .10812,10911. .11018,11125. .11234,
                11323. .11430,11525. .11570,11688. .11819,11866. .12081)
                /note="contains similarity to unknown protein
                gb|AAP44992.1
                gene_id:MVP7.5"
                /codon_start=1
                /evidence=not_experimental
                /protein_id="BAB10306.1"
                /db_xref="GI:10177204"
                /translation="MSATELPTKEAHILKGHEGAVLAARFNGDGNYALTCGKDRTIRL
                WNPHRGILIKTYKSHGREVRDVHVTSDNAKFCSCGGDRQVYYWDVSTGRVIRKFRGHD
                GEVNAVKFNDSSSVVVSAGFDRSLRVWDCRSHSVEPVQIIDTFLDTVMSVVLTKTEII
                GGSVDGTVRTFDMRIGREMSDNLGQPVNCISISNDGNCVLAGCLDSTLRLLDRTTGEL
                LQVYKGHISKSFKTDCCLTNSDAHVIGGSEDGLVFFWDLVDAKVLSKFRAHDLVNSID
                VDFYLRLLKFFVAGDKRELPPEGRLYVDFFGRWYNSCLEKMKKQQILTSIDFVIYVSN
                SPFFFYYRL"
tRNA            complement(12371. .12445)
                /product="tRNA-His"
                /note="codon recognized: CAC; gene_id:MVP7.6"
                /evidence=not_experimental
CDS             join(18535. .18615,19010. .19205,19294. .19496,19580. .19857,
                19940. .20206,20290. .20635,20823. .20960,21047. .21172,
                21284. .21496,21580. .21856,21942. .22138,22221. .22571,
                22732. .23313)
                /note="gene_id:MVP7.7"
                /codon_start=1
                /evidence=not_experimental
                /product="cellulose synthase catalytic subunit"
                /protein_id="BAB10307.1"
                /db_xref="GI:10177205"
                /translation="MNTGGRLIAGSHNRNEFVLINADENARIRSVQELSGQTCQICRD
                EIELTVDGEPFVACNECAFPVCRPCYEYERREGNQACPQCKTRFKRLKGSPRVEGDEE
                EDDIDDVENEFEYGNNGIGFDQVSEGMSISRRNSGFPQSDLDSAPPGSQIPLLTYGDE
                DVEISSLPHALIVPPSLGGHGNRVHPVSLSDPTVAAHPRPMVPQKDLAVYGYGSVAWK
                DRMEEWKRKQNEKLQVVRHEGDPDFEDGDDADFPMMDEGRQPLSRKIPIKSSKINPYR
                MLIVLRLVILGLFFHYRILHPVKDAYALWLISVICEIWFAVSWVLDQFPKWYPIERET
                YLDRLSLRYEKEGKPSGLSPVDVFVSTVDPLKEPPLITANTVLSILAVDYPVDKVACY
                VSDDGAAMLTFEALSETAEFARKWVPFCKKYCIEPRAPEWYFCHKMDYLKNKVHPAFV
                RERRAMKRDYEEFKVKINALVATAQKVPEDGWTMQDGTPWPGNSVRDHPGMIQVFLGS
                DGVRDVENNELPRLVYVSREKRPGFDHHKKAGAMNSLIRVSGVLSNAPYLLNVDCDHY
                INNSKAIREAMCFMMDPQSGKKICYVQFPQRFDGIDRHDRYSNRNVVFFDINMKGLDG
                LQGPIYVGTGCVFRRQALYGFDAPKKKKGPRKTCNCWPKWCLLCFGSRKNRKAKTVAA
                DKKKKNFEASKQIHALENIEEGRVTKGSNVEQSTEAMQMKLEKKFGQSPVFVASARME
                NGGMARNASPACLLKEAIQVISCGYEDKTEWGKEIGWIYGSVTEDILTGFKMHSHGWR
                SVYCTPKLAAFKGSAPINLSDRLHQVLRWALGSVEIFLSRHCPIWYGYGGGLKWLERL
                SYINSVVYPWTSLPLIVYCSLPAICLLTGKFIVPEISNYASILFMALFSSIAITGILE
                MQWGKVGIDDWWRNEQFWVIGGVSAHLFALFQGLLKVLAGVDTNFTVTSKAADDGEFS
                DLYLFKWTSLLIPPMTLLIINVIGVVSGISDAISNGYDSWGPLFGRLFFALWVIIHLY
                PFLKGLMGKQDRMPTIIVVWSILLASIFTLLWVRVNPFVAKGGPILEICGLDCL"
CDS             join(29459. .29897,29970. .30636)
                /note="contains similarity to transcription factor
                gene_id:MVP7.8"
                /codon_start=1
                /evidence=not_experimental
                /protein_id="BAB10308.1"
                /db_xref="GI:10177206"
                /translation="MCVLKVANQEDNVGKKAESIRDDDHRTLSEIDQWLYLFAAEDDH
                HRHSFPTQQPPPSSSSSSLISGFSREMEMSAIVSALTHVVAGNVPQHQQGGGEGSGEG
                TSNSSSSSGQKRRREVEEGGAKAVKAANTLTVDQYFSGGSSTSKVREASSNMSGPGPT
                YEYTTTATASSETSSFSGDQPRRRYRGVRQRPWGKWAAEIRDPFKAARVWLGTFDNAE
                SAARAYDEAALRFRGNKAKLNFPENVKLVRPASTEAQPVHQTAAQRPTQSRNSGSTTT
                LLPIRPASNQSVHSQPLMQSYNLSYSEMARQQQQFQQHHQQSLDLYDQMSFPLRFGHT
                GGSMMQSTSSSSSHSRPLFSPAAVQPPPESASETGYLQDIQWPSDKTSNNYNNSPSS"
CDS             complement(join(31907. .31225,31303. .31380,31479. .31632,
                31734. .31846,31941. .32040,32184. .32380,32485. .32541,
                32640. .32744,32857. .32964,33125. .33319,33423. .33552,
                33700. .33833))
                /note="gene_id:MVP7.9"
                /codon_start=1
                /evidence=not_experimental
                /product="proteasome regulatory subunit-like"
                /protein_id="BAB10309.1"
                /db_xref="GI:10177207"
                /translation="MGLKKLTQMGHMDLRFSFITNGSYSFVHNILPNSYLGMTLQATG
                RGWGFDANILSLSEPLEVKSVSSPESFTNSSSLSRSIDSRRLAMEESRQLESSIDRLL
                NEEKQMRLAENVAGTRKAATEILKLCFEAKDWKLLNEQILNLSKKRGQLKQAVQSMVQ
                QAMQYIDQTLDIETRIELIKTLNNVAAGKIYVEIERARLTKMLAKIKEEQGLIAEAAD
                LMQEVAVETFGAMAKTEKIAFILEQVRLCLDQKDFVRAQILSRKINPRVFDADTTKEK
                KKPKEGENMVEEAPADIPTLLVLKRIYYELMIRYYSHNNEYIEICRSYKAIYDIPSVK
                ENPEQWIPVLRKICWFLALAPHDPMQSSLLNATLEDKKLSEIPDFKMLLKQIVTMEVI
                QWTSLWNKYKDEFENEKNMIGGSLGDKAGEDLKLRIIEHNILVVSKYYSRITFKRLAE
                LLCLTTEEAEKHLSEMVVSKALIAKIDRPSGIICFQIVKDSNEILNSWATNLEKLLDL
                VEKSCHQIHKETMVHKVALRA"
CDS             complement(join(34661. .34824,34990. .35065))
                /note="gene_id:MVP7.10
                unknown protein"
                /codon_start=1
                /evidence=not_experimental
                /protein_id="BAB10310.1"
                /db_xref="GI:10177208"
                /translation="MAIRVSHKSFLVALLLILFISSPTQARSLREVVRNRTLLVVEKS
                QESRKIRHEGGGSDVDGLMDMDYNSANKKRPIHNR"
CDS             complement(join(37680. .37967,38133. .38213,38777. .38935))
                /note="gene_id:MVP7.11
                pir||T04031
                similar to unknown protein"
                /codon_start=1
                /evidence=not_experimental
                /protein_id="BAB10311.1"
                /db_xref="GI:10177209"
```

Query Match 60.9%; Score 2096.4; DB 8; Length 39645;
Best Local Similarity 74.4%; Pred. No. 0;
Matches 3349; Conservative 0; Mismatches 21; Indels 1130; Gaps 11;






b 21103 AGTTTACGTTTCTCGTGAGAAGAGACCCGGATTTGATCACCATAAGAAGGCTGGAGCTAT 21162
y 1618 GAATTCCCT---------- 1626
b 21163 GAATTCCCTGGTAAATGATATACTTTTTAAAGCTCTAAACCTTCTTCTTTGTAAATTACG 21222
y 1627 ---------- 1626
b 21223 TCTTGCCATTTATTGAAATGGTTCCTGACTCTTGATTTCATCTACAAAACTTTTGTTGAA 21282
y 1627 GATACGAGTCTCTGGGGTTCTATCAAATGCTCCTTACCTTCTGAATGTCGATTGTGATCA 1686
b 21283 GATACGAGTCTCTGGGGTTCTATCAAATGCTCCTTACCTTCTGAATGTCGATTGTGATCA 21342
y 1687 CTACATCAACAATAGCAAAGCTCTTAGAGAAGCAATGTGTTTCATGATGGATCCTCAGTC 1746
b 21343 CTACATCAACAATAGCAAAGCTCTTAGAGAAGCAATGTGTTTCATGATGGATCCTCAGTC 21402
y 1747 AGGAAAGAAAATCTGTTATGTTCAGTTCCCTCAAAGGTTCGATGGGATTGATAGGCACGA 1806
b 21403 AGGAAAGAAAATCTGTTATGTTCAGTTCCCTCAAAGGTTCGATGGGATTGATAGGCACGA 21462
y 1807 TCGATACTCAAATCGCAATGTTGTGTTCTTTGAT---------- 1840
b 21463 TCGATACTCAAATCGCAATGTTGTGTTCTTTGATGTAAGTACAGCCACCACTTTCCTATT 21522
y 1841 ----------ATC 1843
b 21523 GTATCCCTTTTTCTTGAGATTTCTGTAGAATACCAACTAATGAATCTTTATTTACAGATC 21582
y 1844 AATATGAAAGGTTTGGATGGGCTACAAGGGCCTATATACGTCGGTACAGGTTGTGTTTTC 1903
b 21583 AATATGAAAGGTTTGGATGGGCTACAAGGGCCTATATACGTCGGTACAGGTTGTGTTTTC 21642
y 1904 AGGAGGCAAGCGCTTTACGGATTTGATGCACCGAAGAAGAAGAAGGGCCCACGTAAGACA 1963
b 21643 AGGAGGCAAGCGCTTTACGGATTTGATGCACCGAAGAAGAAGAAGGGCCCACGTAAGACA 21702
y 1964 TGCAATTGCTGGCCAAAATGGTGTCTCCTATGTTTTGGTTCAAGAAAGAATCGTAAAGCA 2023
b 21703 TGCAATTGCTGGCCAAAATGGTGTCTCCTATGTTTTGGTTCAAGAAAGAATCGTAAAGCA 21762
y 2024 AAGACAGTGGCTGCGGATAAGAAGAAGAAGAATAGGGAAGCGTCAAAGCAGATCCACGCA 2083
b 21763 AAGACAGTGGCTGCGGATAAGAAGAAGAAGAATAGGGAAGCGTCAAAGCAGATCCACGCA 21822
y 2084 TTAGAAAATATCGAAGAGGGCCGCGGTCATAAAGTTCT---------- 2121
b 21823 TTAGAAAATATCGAAGAGGGCCGCGTCACTAAAGGTATCATACAAATCCTGTTTGTTGTT 21882
y 2122 ---------- 2121
b 21883 AAACTCTTTCGTTAGTCGGTGCATTTTACTAAAAAAATAAAATTTAAAAAACATTCTAGG 21942
y 2122 ---TAACGTAGAACAGTCAACCGAGGCAATGCAAATGAAGTTGCAGAAGAAATATGGGCA 2178
b 21943 TTCTAACGTAGAACAGTCAACCGAGGCAATGCAAATGAAGTTGGAGAAGAAATTTGGGCA 22002
y 2179 GTCTCCTGTATTTGTTGCATCTGCGCGTCTGGAGAATGGTGGGATGGCTAGAAACGCAAG 2238
b 22003 GTCTCCTGTATTTGTTGCATCTGCGCGTATGGAGAATGGTGGGATGGCTAGAAACGCAAG 22062
y 2239 CCCGGCTTGTCTGCTTAAAGAAGCCATCCAAGTCATTAGTCGCGGATATGAAGATAAAAC 2298
b 22063 CCCGGCTTGTCTGCTTAAAGAAGCCATCCAAGTCATTAGTTGCGGATATGAAGATAAAAC 22122
y 2299 TGAATGGGGAAAAG---------- 2312
b 22123 TGAATGGGGAAAAGAGGTAAGCAGCCGGTTTTAAACCTTTGTTGTGTTTATTCAATCAAT 22182
y 2313 ----------AGATTGGGTGGATCTATGGTTCTG 2336
b 22183 TCTTGATTTTGATGATGACCTTGTGAAAAAAATCTCAGATTGGGTGGATCTATGGTTCTG 22242


Qy 2337 TTACCGAAGATATT TTACGGGTTCTAAGATGCATTCTCATGGTTGGAGACATGTTTATT 2396
Db 22243 TTACCGAAGATATT TTACGGGTTTTAAGATGCATTCTCATGGTTGGAGATCTGTTTATT 22302
Qy 2397 GTACACCAAAGTTA GGGCTTTCAAAGGATCAGCTCCAATCAATCTTTCGGATCGTCTCC 2456
Db 22303 GTACACCAAAGTTA GGGCTTTCAAAGGATCAGCTCCAATCAATCTTTCGGATCGTCTCC 22362
Qy 2457 ATCAAGTTCTTCGA GGGCGCTTGGGTCGGTTGAGATTTTCTTGAGTAGGCATTGTCCTA 2516
Db 22363 ATCAAGTTCTTCGA GGGCGCTTGGGTCGGTTGAGATTTTCTTGAGTAGGCATTGTCCTA 22422
Qy 2517 TTTGGTATGGTTAT GAGGTGGGTTGAAATGGCTTGAGCGGTTGTCCTACATTAACTCTG 2576
Db 22423 TTTGGTATGGTTAT GAGGTGGGTTGAAATGGCTTGAGCGGTTGTCCTACATTAACTCTG 22482
Qy 2577 TGGTTTACCCGTCT TCTACTATTATA TTACTGTTCTCTCCCTGCCATCTGTC 2636
Db 22483 TGGTTTACCCGTCA TCTACTATTATA TTACTGTTCTCTCCCTGCCATCTGTC 22542
Qy 2637 TTCTCACTGGAAAA TCATCGTTCCCG 2663
Db 22543 TTCTCACTGGAAAA TCATCGTTCCCGAGGTAAAACAATCATCTTGAGTTCTCAAAATAT 22602
Qy 2664 ---------- 2663
Db 22603 GAATCTTTATTTCA GTTTTGTGCTTATTCATTTTCCTTGCCACTGGGGGTTAAAAGTAT 22662
Qy 2664 ---------- 2663
Db 22663 CATATGAATCTTTA TCCAAGTTGTGTGTTTTAAGACCGGAAAACGATTCTTGTTCCTTC 22722
Qy 2664 -------AGATTAG AACTATGCGAGTATCCTCTTCATGGCGCTCTTCTCGTCGATTGCA 2716
Db 22723 TTTTTCCAGATTAG AACTATGCGAGTATCCTCTTCATGGCGCTCTTCTCGTCGATTGCA 22782
Qy 2717 ATAACGGGTATTCT GAGATGCAATGGGGCAAAGTTGGGATCGATGATTGGTGGAGAAAC 2776
Db 22783 ATAACGGGTATTCT GAGATGCAATGGGGCAAAGTTGGGATCGATGATTGGTGGAGAAAC 22842
Qy 2777 GAACAGTTTTGGGT ATTGGAGGTGTTTCTGCGCATCTGTTTGCTCTCTTCCAAGGTCTC 2836
Db 22843 GAACAGTTTTGGGT ATTGGAGGTGTTTCTGCGCATCTGTTTGCTCTCTTCCAAGGTCTC 22902
Qy 2837 CTCAAGGTTCTTGCTGGTGTCGACACTAACTTCACAGTCACATCAAAAGCAGCTGATGAT 2896
Db 22903 CTCAAGGTTCTTGCTGGTGTCGACACTAACTTCACAGTCACATCAAAAGCAGCTGATGAT 22962
Qy 2897 GGAGAGTTCTCTGA CTTTACCTCTTCAAATGGACTTCACTTCTCATCCCTCCAATGACT 2956
Db 22963 GGAGAGTTCTCTGA CTTTACCTCTTCAAATGGACTTCACTTCTCATCCCTCCAATGACT 23022
Qy 2957 CTACTCATCATAAA GTCATTGGAGTCATAGTCGGAGTCTCTGATGCCATCAGCAATGGA 3016
Db 23023 CTACTCATCATAAA GTCATTGGAGTCATAGTCGGAGTCTCTGATGCCATCAGCAATGGA 23082
Qy 3017 TACGACTCGTGGGGACCGCTTTTCGGAAGACTGTTCTTTGCACTTTGGGTCATCATTCAT 3076
Db 23083 TACGACTCGTGGGGACCGCTTTTCGGAAGACTGTTCTTTGCACTTTGGGTCATCATTCAT 23142
Qy 3077 CTTTACCCGTTCCTTAAAGGTTTGCTTGGGAAACAAGATAGAATGCCAACCATTATTGTC 3136
Db 23143 CTTTACCCGTTCCTTAAAGGTTTGCTTGGGAAACAAGATAGAATGCCAACCATTATTGTC 23202
Qy 3137 GTCTGGTCCATCCTCCTGGCCTCGATTCTTACACTTCTTTGGGTCCGGGTTAATCCGTTT 3196
Db 23203 GTCTGGTCCATCCTCCTGGCCTCGATTCTTACACTTCTTTGGGTCCGGGTTAATCCGTTT 23262
Qy 3197 GTGGCGAAAGGCGGTCCTATTCTCGAGATCTGTGGTTTAGACTGCTTGTGATTCGATTGA 3256
Db 23263 GTGGCGAAAGGCGGTCCTATTCTCGAGATCTGTGGTTTAGACTGCTTGTGATTCGATTGA 23322

```
Qy       3257 CCGGTGGATGGGTTGGTGAAAAAGGTTTAATTCCCACGGATCAAAGAGAGGTAAGAGAGA 3316
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      23323 CCGGTGGATGGGTTGGTGAAAAAGGTTTAATTCCCACGGATCAAAGAGAGGTAAGAGAGA 23382

Qy       3317 TATTGTTTTACCTCTAAAAGACTCCTTCATTGTGTTCATTAGATGATGAAAAATGAAAAG 3376
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      23383 TATTGTTTTACCTCTAAAAGACTCCTTCATTGTGTTCATTAGATGATGAAAAATGAAAAG 23442

Qy       3377 AAAAAGAAGATTTAATTTTGTTACGAGAATTGTTATTTTTGCAAGAATGTGTTGTAGATA 3436
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      23443 AAAAAGAAGATTTAATTTTGTTACGAGAATTGTTATTTTTGCAAGAATGTGTTGTAGATA 23502


RESULT 4
AX505864
LOCUS       AX505864               3255 bp    DNA     linear   PAT 27-SEP-2002
DEFINITION  Sequence 559 from Patent WO0216655.
ACCESSION   AX505864
VERSION     AX505864.1  GI:23387101
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Harper,J.F., Kreps,J., Wang,X. and Zhu,T.
  TITLE     Stress-regulated genes of plants, transgenic plants containing
            same, and methods of use
  JOURNAL   Patent: WO 0216655-A 559 28-FEB-2002;
            The Scripps Research Institute (US) ; Syngenta Participations AG
            (CH)
FEATURES             Location/Qualifiers
     source          1..3255
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3702"
BASE COUNT      884 a    621 c    798 g    952 t
ORIGIN

  Query Match           56.4%;  Score 1944;  DB 6;  Length 3255;
  Best Local Similarity  76.3%;  Pred. No. 0;
  Matches 2484;  Conservative    0;  Mismatches  725;  Indels   45;  Gaps    6;

Qy          8 CGGTTAATCGCCGGTTCTCACAACAGGAATGAGTTTGTCCTCATTAATGCCGATGAGAAT 67
              ||| | || || || |||||||||| || || || || || |||||| |||||||||| |
Db         16 CGGCTCATTGCTGGCTCTCACAACAGAAACGAATTCGTTCTCATTAACGCCGATGAGAGT 75

Qy         68 GCCCGAATAAGATCAGTCCAAGAGCTGAGTGGACAGACATGCCAAATCTGCAGAGATGAG 127
              || |||| |||||| || | |||||| || ||| |||||||| |||||||
Db         76 GCCAGAATACGATCAGTACAAGAACTGAGTGGGCAAACATGCCAAATCTGTGGAGATGAA 135

Qy        128 ATCGAATTGACTGTTGATGGAGAACCGTTTGTGGCATGTAAGGAATGTGCATTCCCTGTG 187
              |||||||| || |||    | || | |||| || || |||||| |||||||| ||
Db        136 ATCGAATTAACGGTTAGCAGTGAGCTCTTTGTTGCTTGCAACGAATGCGCATTCCCGGTT 195

Qy        188 TGTAGACCTTGCTATGAGTACGAAAGACGAGAAGGCAATCAAGCTTGTCCACAGTGCAAA 247
              |||||||| |||||||||| || | ||||||| |||||||||||| |||||||||
Db        196 TGTAGACCATGCTATGAGTATGAACGTAGAGAAGGAAATCAAGCTTGTCCTCAGTGCAAA 255

Qy        248 ACCCGTTTCAAACGTCTTAAAGGAAGTCCAAGAGTTGA---AGGTGATGAAGAGGAAGAT 304
              || || | |||| |  ||||||| |||||| | |||||   || ||||||||| |||||
Db        256 ACTCGATACAAAAGGATTAAAGGTAGTCCACGGGTTGATGGAGATGATGAAGAAGAAGAA 315

Qy        305 GACATTGATGATTTAGACAATGAGTTTGAGTATGGAAATAATGGGATTGGATTTGATCAG 364
              |||||||||||| | ||  |||||||||          || | ||||| |    ||| ||
Db        316 GACATTGATGATCTTGAGTATGAGTTTG---------ATCATGGGATGGACCCTGAACAT 366

Qy        365 GTTTCTGAAGGTATGTCAATCTCTCGTCGCAACTCCGGTTTCCCACAATCTGATTTGGAT 424
              |   |||||           || | |  ||| |||||      |    ||  ||||||
Db        367 GCCGCTGAAGCCGCACTCTCTTCACGCCTTAACACCGGT------CGTGGTGGATTGGAT 420

Qy        425 TCAGCTCCACCTGGCTCTCAGATTCCATTGCTGACTTACGGCGACGAGGACGTTGAGATT 484
              |||||||||||||||||||||||||  | |||||||  | || || || | ||| ||
Db        421 TCAGCTCCACCTGGCTCTCAGATTCCTCTTTTGACTTATTGTGATGAAGATGCTGATATG 480

Qy        485 TCTTCTGATAGACATGCTCTTATTGTTCCTCCTTCACTTGGTGGTCATGGCAATAGAGTT 544
              | ||||||| | |||||||||||| || |||||||||      ||  |||| ||| | ||
Db        481 TATTCTGATCGTCATGCTCTTATCGTGCCTCCTTCA---ACGGGATATGGGAATCGCGTC 537

Qy        545 CATCCTGTTTCTCTTTCTGACCCGACCGTGGCTGCACATCGAAGGCTGATGGTACCTCAG 604
              ||||||   |  || | ||  |  | |   || ||||    ||    ||||| ||||||
Db        538 TATCCTGCACCGTTTACAGATTCTTCTGCACCTCCACAGGCGAGATCAATGGTTCCTCAG 597

Qy        605 AAAGATCTTGCGGTTTATGGTTATGGAAGTGTCGCTTGGAAAGATCGGATGGAGGAATGG 664
              |||||| ||||||  |||||||||||||||||| ||||||||  || ||  |||||  |  |||
Db        598 AAAGATATTGCGGAATATGGTTATGGAAGTGTTGCTTGGAAGGACCGTATGGAAGTTTGG 657

Qy        665 AAGAGAAAGCAGAATGAGAAACTTCAGGTTGTTAGGCATGAAGGAG-------------- 710
              ||||||   ||    || || ||||| ||  ||| ||||||||||
Db        658 AAGAGACGACAAGGCGAAAAGCTTCAAGTCATTAAGCATGAAGGAGGAAACAATGGTCGA 717

Qy        711 ----ATCCTGATTTTGAAGATGGTGATGATGCTGATTTTCCAATGATGGATGAGGGAAGG 766
                  ||||   || ||    |||||| ||||  | || ||||||||||| |||||
Db        718 GGTTCCAATGATGACGACGAACTAGATGATCCTGACATGCCTATGATGGATGAAGGAAGA 777

Qy        767 CAGCCATTGTCTATGAAGATACCAATCAAATCGAGCAAGATAAATCCTTACCGGATGTTA 826
              || ||  | || |  ||| |||| ||    || |||| ||||||||||| ||||||||
Db        778 CAACCTCTCTCAAGAAAGCTACCTATTCGTTCAAGCAGAATAAATCCTTACAGGATGTTA 837

Qy        827 ATTGTGCTACGTCTTGTGATTCTTGGTCTCTTCTTTCACTACCGTATTCTTCACCCCGTC 886
              ||| ||   || || | |||||||||||| |||||||| ||  | |||| || || |||
Db        838 ATTCTGTGTCGCCTCGCGATTCTTGGTCTTTTCTTTCATTATAGAATTCTCCATCCAGTC 897

Qy        887 AAAGATGCATATGCTTTGTGGCTTATTTCTGTTATATGTGAGATATGGTTTGCTGTTTCA 946
              || ||||||||||  || ||| | |  || |||||||| |||||||||||||| || ||
Db        898 AATGATGCATATGGATTATGGTTAACGTCAGTTATATGCGAGATATGGTTTGCAGTGTCT 957

Qy        947 TGGGTTCTTGATCAGTTCCCTAAATGGTACCCTATCGAGCGAGAAACGTACTTGGACCGA 1006
              ||| ||||||||| ||||| |||||||| ||||| || || ||||| ||| | ||  ||
Db        958 TGGATTCTTGATCAATTCCCCAAATGGTATCCTATAGAACGTGAAACATACCTCGATAGA 1017

Qy       1007 CTCTCATTAAGATATGAGAAAGAAGGGAAACCGTCGGGACTATCCCCTGTGGATGTATTT 1066
              |||||  | || || ||||| ||||| |||||||| ||| || | ||||| ||||| |||
Db       1018 CTCTCTCTCAGGTACGAGAAGGAAGGAAAACCGTCAGGATTAGCACCTGTTGATGTTTTT 1077

Qy       1067 GTTAGTACAGTGGATCCATTGAAAGAGCCTCCGCTTATTACTGCAAATACTGTCTTGTCT 1126
              |||||||||||||||| |||||||||| ||  | |||| |||| || || | ||
Db       1078 GTTAGTACAGTGGATCCGTTGAAAGAGCCACCCTTGATTACAGCAAACACAGTTCTTTCC 1137

Qy       1127 ATTCTTGCTGTTGATTATCCTGTCGATAAGGTTGCTTGTTACGTATCTGATGATGGTGCT 1186
              ||||| || |||||||||||| |||||||||| ||||| ||||| || |||||||||
Db       1138 ATTCTAGCAGTTGATTATCCTGTGGATAAGGTTGCGTGTTATGTATCAGACGATGGTGCA 1197

Qy       1187 GCTATGCTTACTTTCGAAGCTCTTTCTGAGACCGCTGAATTCGCAAGGAAATGGGTTCCT 1246
              |||||||||| || |||||||| ||||| || |||| || || || ||||||||||||
Db       1198 GCTATGCTTACATTTGAAGCTCTCTCTGATACAGCTGAGTTTGCTAGAAAATGGGTTCCT 1257

Qy       1247 TTCTGCAAGAAATATTGTATTGAGCCTCGTGCTCCCGAATGGTATTTCTGCCATAAAATG 1306
              || || ||||| | |  ||| ||||| || ||||| || |||||||| |  || || |||
Db       1258 TTTTGTAAGAAGTTTAATATCGAGCCACGAGCTCCTGAGTGGTATTTTTCTCAGAAGATG 1317

Qy       1307 GACTACTTGAAGAATAAAGTTCATCCCGCATTTGTTAGGGAGCGGCGAGCCATGAAGAGA 1366
              || ||| ||||||| |||||||||| || ||||| ||||| || || |||||||||||
Db       1318 GATTACCTGAAGAACAAAGTTCATCCTGCTTTTGTCAGGGAACGTCGTGCTATGAAGAGA 1377

Qy       1367 GATTATGAAGAATTCAAAGTAAAGATCAATGCTTTAGTAGCAACAGCACAGAAAGTGCCT 1426
              |||||||| || || ||||| ||||| |||||   | || || || ||||||||||||||
Db       1378 GATTATGAGGAGTTTAAAGTGAAGATAAATGCACTGGTTGCTACTGCACAGAAAGTGCCT 1437

Qy       1427 GAGGATGGTTGGACTATGCAAGACGGTACACCTTGGCCCGGTAATAGTGTGCGAGATCAT 1486
              ||||| |||||||||||||||| || || |||||||| || || |  || || || |||
Db       1438 GAGGAAGGTTGGACTATGCAAGATGGAACTCCTTGGCCTGGAAACAACGTCCGTGACCAT 1497

Qy       1487 CCTGGCATGATTCAGGTCTTCCTTGGAAGTGACGGTGTTCGTGATGTCGAAAACAACGAG 1546
              ||||| |||||||||||| ||| | ||   |  || |||||||||   ||    || |||
```

```
b    1498 CCTGGAATGATTCAGGTGTTCTTGGGTCATAGTGGAGTTCGTGATACGGATGGTAATGAG 1557
y    1547 TTGCCTCGATTAGTTTACGTTTCTCGTGAGAAGAGACCCGGATTTGATCACCATAAGAAG 1606
b    1558 TTACCACGTCTAGTGTATGTTTCTCGTGAGAAGCGGCCTGGATTTGATCACCACAAGAAA 1617
y    1607 GCTGGAGCTATGAATTCCCTGATACGAGTCTCTGGGGTTCTATCAAATGCTCCTTACCTT 1666
b    1618 GCTGGAGCTATGAATTCCTTGATCCGAGTCTCTGCTGTTCTATCAAACGCTCCTTACCTT 1677
y    1667 CTGAATGTCGATTGTGATCACTACATCAACAATAGCAAAGCTCTTAGAGAAGCAATGTGT 1726
b    1678 CTTAATGTCGATTGTGATCACTACATCAACAACAGCAAAGCAATTAGAGAATCTATGTGT 1737
y    1727 TTCATGATGGATCCTCAGTCAGGAAAGAAAATCTGTTATGTTCAGTTCCCTCAAAGGTTC 1786
b    1738 TTCATGATGGACCCGCAATCGGGAAAGAAAGTTTGTTATGTTCAGTTTCCGCAGAGATTT 1797
y    1787 GATGGGATTGATAGGCACGATCGATACTCAAATCGCAATGTTGTGTTCTTTGATATCAAT 1846
b    1798 GATGGGATTGATAGACATGATAGATACTCAAACCGTAACGTTGTGTTCTTTGATATTAAC 1857
y    1847 ATGAAAGGTTTGGATGGGCTACAAGGGCCTATATACGTCGGTACAGGTTGTGTTTTCAGG 1906
b    1858 ATGAAAGGTCTTGATGGGATACAAGGACCGATATATGTCGGGACAGGTTGTGTGTTTAGA 1917
y    1907 AGGCAAGCGCTTTACGGATTTGATGCACCGAAGAAGAAGAAGGGCCCACGTAAGACATGC 1966
b    1918 AGACAGGCTCTTTATGGTTTTGATGCACCAAAGAAGAAGAAACCACCAGGCAAAACCTGT 1977
y    1967 AATTGCTGGCCAAAATGGTGTCTCCTATGTTTTGGTTCAAGAAAGAATCGTAAAGCAAAG 2026
b    1978 AACTGTTGGCCTAAATGGTGTTGTTTGTGTTGTGGGTTGAGAAAGAAGAGTAAAACGAAA 2037
y    2027 ACAGTGGCTGCGGATAAGAAGAAGAAGAATAGGGAAGCGTCAAAGCAGATCCACGCATTA 2086
b    2038 ------GCCAAAGATAAGAAAACTAACACTAAAGAGACTTCAAAGCAGATTCATGCGCTA 2091
y    2087 GAAAATATCGAAGAGGGCCGCGGTCATAAAGTTCTTAACGTAGAACAGTCAACCGAGGCA 2146
b    2092 GAGAATGTCGACGAAGGTGTTATCGTCCCAGTGTCAAATGTTGAGAAGAGATCTGAAGCA 2151
y    2147 ATGCAAATGAAGTTGCAGAAGAAATATGGGCAGTCTCCTGTATTTGTTGCATCTGCGCGT 2206
b    2152 ACACAATTGAAATTGGAGAAGAAGTTTGGACAATCTCCGGTTTTCGTTGCCTCTGCTGTT 2211
y    2207 CTGGAGAATGGTGGGATGGCTAGAAACGCAAGCCCGGCTTGTCTGCTTAAAGAAGCCATC 2266
b    2212 CTACAGAACGGTGGAGTTCCCCGTAACGCAAGCCCCGCATGTTTGTTAAGAGAAGCCATT 2271
y    2267 CAAGTCATTAGTCGCGGATATGAAGATAAAACTGAATGGGGAAAAGAGATTGGGTGGATC 2326
b    2272 CAAGTTATTAGCTGCGGGTACGAAGATAAAACCGAATGGGGAAAAGAGATCGGGTGGATT 2331
y    2327 TATGGTTCTGTTACCGAAGATATTCTTACGGGTTCTAAGATGCATTCTCATGGTTGGAGA 2386
b    2332 TATGGATCGGTGACTGAAGATATCCTGACGGGTTTCAAGATGCATTGCCATGGATGGAGA 2391
y    2387 CATGTTTATTGTACACCAAAGTTAGCGGCTTTCAAAGGATCAGCTCCAATCAATCTTTCG 2446
b    2392 TCTGTGTACTGTATGCCTAAGCGTGCAGCTTTTAAAGGATCTGCTCCTATTAACTTGTCA 2451
y    2447 GATCGTCTCCATCAAGTTCTTCGATGGGCGCTTGGGTCGGTTGAGATTTTCTTGAGTAGG 2506
b    2452 GATCGTCTTCATCAAGTTCTACGTTGGGCTCTTGGCTCTGTAGAGATTTTCTTGAGCAGA 2511
y    2507 CATTGTCCTATTTGGTATGGTTATGGAGGTGGGTTGAAATGGCTTGAGCGGTTGTCCTAC 2566
b    2512 CATTGTCCGATATGGTATGGTTATGGTGGTGGTTTAAAATGGTTGGAGAGATTCTCTTAC 2571
y    2567 ATTAACTCTGTGGTTTACCCGTGGACCTCTCTACCGCTCATCGTTTACTGTTCTCTCCCT 2626
b    2572 ATCAACTCTGTCGTCTATCCTTGGACTTCACTTCCATTGATCGTCTATTGTTCTCTCCCC 2631

Qy   2627 GCCATCTGTCTTCTCACTGGAAAATTCATCGTTCCCGAGATTAGCAACTATGCGAGTATC 2686
Db   2632 GCGGTTTGTTTACTCACAGGAAAATTCATCGTCCCTGAGATAAGCAACTACGCAGGTATA 2691
Qy   2687 CTCTTCATGGCGCTCTTCTCGTCGATTGCAATAACGGGTATTCTCGAGATGCAATGGGGC 2746
Db   2692 CTCTTCATGCTCATGTTCATATCCATAGCAGTAACTGGAATCCTCGAAATGCAATGGGGA 2751
Qy   2747 AAAGTTGGGATCGATGATTGGTGGAGAAACGAACAGTTTTGGGTCATTGGAGGTGTTTCT 2806
Db   2752 GGTGTCGGAATCGATGATTGGTGGAGAAACGAGCAGTTTTGGGTAATCGGAGGGGCCTCC 2811
Qy   2807 GCGCATCTGTTTGCTCTCTTCCAAGGTCTCTCAAGGTTCTTGCTGGTGTCGACACTAAC 2866
Db   2812 TCGCATCTATTTGCTCTGTTTCAAGGTTTGCTCAAAGTTCTAGCCGGAGTTAACACGAAT 2871
Qy   2867 TTCACAGTCACATCAAAGCAGCTGATGATGAAGTTCTCTGACCTTTACCTCTTCAAA 2926
Db   2872 TTCACAGTCACTTCAAAGCAGCAGACGATGAAGTTTCTCTGAGCTTTACATCTTCAAG 2931
Qy   2927 TGGACTTCACTTCTCATCCCTCCAATGACTCTACTCATCATAAACGTCATTGGAGTCATA 2986
Db   2932 TGGACAACTTTGTTGATTCCTCCGACAACACTTCTGATCATTAACATCATTGGAGTTATT 2991
Qy   2987 GTCGGAGTCTCTGATGCCATCAGCAATGGATACGACTCGTGGGGACCGCTTTTCGGAAGA 3046
Db   2992 GTCGGCGTTTCTGATGCCATTAGCAATGGCTATGACTCATGGGGACCTCTCTTTGGGAGA 3051
Qy   3047 CTGTTCTTTGCACTTTGGGTCATCATTCATCTTTACCCGTTCCTTAAAGGTTTGCTTGGG 3106
Db   3052 CTTTTCTTCGCTCTTTGGGTCATTGTTCATTTATACCCATTCCTCAAGGGAATGCTTGGG 3111
Qy   3107 AAACAAGATAGAATGCCAACCATTATTGTCGTCTGGTCCATCCTCCTGGCCTCGATTCTT 3166
Db   3112 AAGCAAGACAAAATGCCTACGATTATTGTGGTCTGGTCTATTCTTCTAGCTTCGATCTTG 3171
Qy   3167 ACACTTCTTTGGGTCCGGGTTAATCCGTTTGTGGCGAAAGGCGGTCCTATTCTCGAGATC 3226
Db   3172 ACACTCTTGTGGGTCAGAGTTAACCCGTTTGTGGCTAAAGGGGGACCAGTGTTGGAGATC 3231
Qy   3227 TGTGGTTTAGACTG 3240
Db   3232 TGTGGTCTGAATTG 3245
```

```
RESULT 5
AF027173
LOCUS       AF027173               3771 bp    mRNA    linear   PLN 07-FEB-1998
DEFINITION  Arabidopsis thaliana cellulose synthase catalytic subunit (Ath-A)
            mRNA, complete cds.
ACCESSION   AF027173
VERSION     AF027173.1  GI:2827140
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 3771)
  AUTHORS   Arioli,T., Peng,L., Betzner,A.S., Burn,J., Wittke,W., Herth,W.,
            Camilleri,C., Hofte,H., Plazinski,J., Birch,R., Cork,A., Glover,J.,
            Redmond,J. and Williamson,R.E.
  TITLE     Molecular analysis of cellulose biosynthesis in Arabidopsis
  JOURNAL   Science 279 (5351), 717-720 (1998)
  MEDLINE   98111412
   PUBMED   9445479
REFERENCE   2  (bases 1 to 3771)
  AUTHORS   Arioli,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-SEP-1997) Plant Science Centre, Australian National
            University, Acton, Canberra, ACT 200, Australia
FEATURES             Location/Qualifiers
```

```
     source          1. .3771
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /cultivar="Columbia"
                     /db_xref="taxon:3702"
     gene            1. .3771
                     /gene="Ath-A"
     CDS             217. .3471
                     /gene="Ath-A"
                     /note="RSW1-like"
                     /codon_start=1
                     /product="cellulose synthase catalytic subunit"
                     /protein_id="AAC39335.1"
                     /db_xref="GI:2827141"
                     /translation="MNTGGRLIAGSHNRNEFVLINADESARIRSVQELSGQTCQICGD
                     EIELTVSSELFVACNECAFPVCRPCYEYERREGNQACPQCKTRYKRIKGSPRVDGDDE
                     EEEDIDDLEYEFDHGMDPEHAAEAALSSRLNTGRGGLDSAPPGSQIPLLTYCDEDADM
                     YSDRHALIVPPSTGYGNRVYPAPFTDSSAPPQARSMVPQKDIAEYGYGSVAWKDRMEV
                     WKRRQGEKLQVIKHEGGNNGRGSNDDDELDDPDMPMMDEGRQPLSRKLPIRSSRINPY
                     RMLILCRLAILGLFFHYRILHPVNDAYGLWLTSVICEIWFAVSWILDQFPKWYPIERE
                     TYLDRLSLRYEKEGKPSGLAPVDVFVSTVDPLKEPPLITANTVLSILAVDYPVDKVAC
                     YVSDDGAAMLTFEALSDTAEFARKWVPFCKKFNIEPRAPEWYFSQKMDYLKNKVHPAF
                     VRERRAMKRDYEEFKVKINALVATAQKVPEEGWTMQDGTPWPGNNVRDHPGMIQVFLG
                     HSGVRDTDGNELPRLVYVSREKRPGFDHHKKAGAMNSLIRVSAVLSNAPYLLNVDCDH
                     YINNSKAIRESMCFMMDPQSGKKVCYVQFPQRFDGIDRHDRYSNRNVVFFDINMKGLD
                     GIQGPIYVGTGCVFRRQALYGFDAPKKKKPPGKTCNCWPKWCCLCCGLRKKSKTKAKD
                     KKTNTKETSKQIHALENVDEGVIVPVSNVEKRSEATQLKLEKKFGQSPVFVASAVLQN
                     GGVPRNASPACLLREAIQVISCGYEDKTEWGKEIGWIYGSVTEDILTGFKMHCHGWRS
                     VYCMPKRAAFKGSAPINLSDRLHQVLRWALGSVEIFLSRHCPIWYGYGGGLKWLERFS
                     YINSVVYPWTSLPLIVYCSLPAVCLLTGKFIVPEISNYAGILFMLMFISIAVTGILEM
                     QWGGVGIDDWWRNEQFWVIGGASSHLFALFQGLLKVLAGVNTNFTVTSKAADDGAFSE
                     LYIFKWTTLLIPPTTLLIINIIGVIVGVSDAISNGYDSWGPLFGRLFFALWVIVHLYP
                     FLKGMLGKQDKMPTIIVVWSILLASILTLLWVRVNPFVAKGGPVLEICGLNCGN"
BASE COUNT      1037 a    699 c    907 g   1128 t
ORIGIN
```

```
  Query Match           56.4%;  Score 1944;  DB 8;  Length 3771;
  Best Local Similarity 76.3%;  Pred. No. 0;
  Matches 2484;  Conservative    0;  Mismatches  725;  Indels   45;  Gaps    6;

Qy          8 CGGTTAATCGCCGGTTCTCACAACAGGAATGAGTTTGTCCTCATTAATGCCGATGAGAAT 67
              ||| | || || || |||||||||||| || || || |||||||| ||||||||||
Db        232 CGGCTCATTGCTGGCTCTCACAACAGAAACGAATTCGTTCTCATTAACGCCGATGAGAGT 291

Qy         68 GCCCGAATAAGATCAGTCCAAGAGCTGAGTGGACAGACATGTCAAATCTGCAGAGATGAG 127
              ||| |||| |||||||  |||| |||||||  || ||||||||||  |||||||
Db        292 GCCAGAATACGATCAGTACAAGAACTGAGTGGCCAAACATCCAAATCTGTGGAGATGAA 351

Qy        128 ATCGAATTGACTGTTGATGGAGAACGGTTTGTGGCATGTAAGAATGTGCATTCCCTGTG 187
              |||.|||| || |||    | || |  |||.| |. || ||  |||| ||||||||  ||
Db        352 ATCGAATTAACGGTTAGCAGTGAGCTCTTTGTTGCTTGCAAGGAATGCGCATTCCCGGTT 411

Qy        188 TGTAGACCTTGCTATGAGTACGAAAGACGAGAAGGCAATCAAGCTTGTCCACAGTGCAAA 247
              |||||||| |||||||||||| ||| |  |||||||  |||||||||||||||||||||
Db        412 TGTAGACCATGCTATGAGTATGAACGTAGAGAAGGAAATCAAGCTTGTCCTCAGTGCAAA 471

Qy        248 ACCCGTTTCAAACGTCTTAAAGGAAGTCCAAGAGTTGA--AGGTGATGAAGAGGAAGAT 304
              || || | |||| |  |||||||  |||'|| | ||||   || |||||||| |||||
Db        472 ACTCGATACAAAAGGATTAAAGGTAGTCCACGGGTTGATGGAGATGATGAAGAAGAAGAA 531

Qy        305 GACATTGATGATTTAGACAATGAGTTTGAGTATGGAAATAATGGGATTGGATTTGATCAG 364
              |||||||||||| | ||  |||||||||          || ||.|||  |    ||| ||
Db        532 GACATTGATGATCTTGAGTATGAGTTTG----  ----ATCATGGGATGGACCCTGAACAT 582

Qy        365 GTTTCTGAAGGTATGTCAATCTCTCGTCGCAACTCCGGTTTCCCACAATCTGATTTGGAT 424
              |   ||||||           || || |  ||| ||||      ||  ||||||
Db        583 GCCGCTGAAGCCGCACTCTCTTCACGCCTTAACACCGGT-   ---CGTGGTGGATTGGAT 636

Qy        425 TCAGCTCCACCTGGCTCTCAGATTCCATTGCTGACTTACGGCGACGAGGACGTTGAGATT 484
              |||||||||||||||||||||||||||  |  |||||||  | || || || | ||| ||
Db        637 TCAGCTCCACCTGGCTCTCAGATTCCTCTTTTGACTTATTGTGATGAAGATGCTGATATG 696

Qy        485 TCTTCTGATAGACATGCTCTTATTGTTCCTCCTTCACTTGGTGGTCATGGCAATAGAGTT 544
```

```
Db        697 TATTCTGATCGTCATGCTCTTATCGTGCCTCCTTCA---ACGGGATATGGGAATCGCGTC 753

Qy        545 CATCCTGTTTCTCTTTCTGACCCGACCGTGGCTGCACATCGAAGGCTGATGGTACCTCAG 604
              ||||||   | || | ||  |  | |  || ||||    ||    ||||| ||||||
Db        754 TATCCTGCACCGTTTACAGATTCTTCTGCACCTCCACAGGCGAGATCAATGGTTCCTCAG 813

Qy        605 AAAGATCTTGCGGTTTATGGTTATGGAAGTGTCGCTTGGAAAGATCGGATGGAGGAATGG 664
              |||||| ||||||  |||||||||||||||||| |||||||| || || ||||| |  |||
Db        814 AAAGATATTGCGGAATATGGTTATGGAAGTGTTGCTTGGAAGGACCGTATGGAAGTTTGG 873

Qy        665 AAGAGAAAGCAGAATGAGAAACTTCAGGTTGTTAGGCATGAAGGAG-------------- 710
              ||||||   ||    || || ||||| ||  ||| ||||||||||
Db        874 AAGAGACGACAAGGCGAAAAGCTTCAAGTCATTAAGCATGAAGGAGGAAACAATGGTCGA 933

Qy        711 ----ATCCTGATTTTGAAGATGGTGATGATGCTGATTTTCCAATGATGGATGAGGGAAGG 766
                  ||||   || ||    |||||| |||| | || |||||||||| |||||
Db        934 GGTTCCAATGATGACGACGAACTAGATGATCCTGACATGCCTATGATGGATGAAGGAAGA 993

Qy        767 CAGCCATTGTCTATGAAGATACCAATCAAATCGAGCAAGATAAATCCTTACCGGATGTTA 826
              || ||  | || |  ||| |||| ||   || ||||  |||||||||||| ||||||||
Db        994 CAACCTCTCTCAAGAAAGCTACCTATTCGTTCAAGCAGAATAAATCCTTACAGGATGTTA 1053

Qy        827 ATTGTGCTACGTCTTGTGATTCTTGGTCTCTTCTTTCACTACCGTATTCTTCACCCCGTC 886
              ||| ||   || || | |||||||||||| |||||||| ||  | |||| || || |||
Db       1054 ATTCTGTGTCGCCTCGCGATTCTTGGTCTTTTCTTTCATTATAGAATTCTCCATCCAGTC 1113

Qy        887 AAAGATGCATATGCTTTGTGGCTTATTTCTGTTATATGTGAGATATGGTTTGCTGTTTCA 946
              || |||||||||  || ||| | |  || |||||||| |||||||||||||| || ||
Db       1114 AATGATGCATATGGATTATGGTTAACGTCAGTTATATGCGAGATATGGTTTGCAGTGTCT 1173

Qy        947 TGGGTTCTTGATCAGTTCCCTAAATGGTACCCTATCGAGCGAGAAACGTACTTGGACCGA 1006
              ||| |||||||| |||| |||||||| || || |||| ||| | ||  ||
Db       1174 TGGATTCTTGATCAATTCCCCAAATGGTATCCTATAGAACGTGAAACATACCTCGATAGA 1233

Qy       1007 CTCTCATTAAGATATGAGAAAGAAGGGAAACCGTCGGGACTATCCCCTGTGGATGTATTT 1066
              |||||  | || || ||||| ||||| |||||||| ||| || | |||| ||||| |||
Db       1234 CTCTCTCTCAGGTACGAGAAGGAAGGAAAACCGTCAGGATTAGCACCTGTTGATGTTTTT 1293

Qy       1067 GTTAGTACAGTGGATCCATTGAAAGAGCCTCCGCTTATTACTGCAAATACTGTCTTGTCT 1126
              |||||||||||||||| |||||||||| || | |||| |||| || || | ||
Db       1294 GTTAGTACAGTGGATCCGTTGAAAGAGCCACCCTTGATTACAGCAAACACAGTTCTTTCC 1353

Qy       1127 ATTCTTGCTGTTGATTATCCTGTCGATAAGGTTGCTTGTTACGTATCTGATGATGGTGCT 1186
              ||||| || |||||||||||| ||||| |||| || |||||||
Db       1354 ATTCTAGCAGTTGATTATCCTGTGGATAAGGTTGCGTGTTATGTATCAGACGATGGTGCA 1413

Qy       1187 GCTATGCTTACTTTCGAAGCTCTTTCTGAGACCGCTGAATTCGCAAGGAAATGGGTTCCT 1246
              |||||||||| || |||||||| ||||| || ||||| || || || ||||||||||||
Db       1414 GCTATGCTTACATTTGAAGCTCTCTCTGATACAGCTGAGTTTGCTAGAAAATGGGTTCCT 1473

Qy       1247 TTCTGCAAGAAATATTGTATTGAGCCTCGTGCTCCCGAATGGTATTTCTGCCATAAAATG 1306
              || || ||||| | |  ||| ||||| || ||||| || |||||||| |  || || |||
Db       1474 TTTTGTAAGAAGTTTAATATCGAGCCACGAGCTCCTGAGTGGTATTTTTCTCAGAAGATG 1533

Qy       1307 GACTACTTGAAGAATAAAGTTCATCCCGCATTTGTTAGGGAGCGGCGAGCCATGAAGAGA 1366
              || ||| |||||| |||||||||| || ||||| ||||| || || || ||||||||||
Db       1534 GATTACCTGAAGAACAAAGTTCATCCTGCTTTGTCAGGGAACGTCGTGCTATGAAGAGA 1593

Qy       1367 GATTATGAAGAATTCAAAGTAAAGATCAATGCTTTAGTAGCAACAGCACAGAAAGTGCCT 1426
              |||||||| || || ||||| ||||| |||||  | || || || ||||||||||||||||
Db       1594 GATTATGAGGAGTTTAAAGTGAAGATAAATGCACTGGTTGCTACTGCACAGAAAGTGCCT 1653

Qy       1427 GAGGATGGTTGGACTATGCAAGACGGTACACCTTGGCCCGGTAATAGTGTGCGAGATCAT 1486
              ||||| |||||||||||||||||| || |||||||| || || |  || || || |||
Db       1654 GAGGAAGGTTGGACTATGCAAGATGGAACTCCTTGGCCTGGAAACAACGTCCGTGACCAT 1713

Qy       1487 CCTGGCATGATTCAGGTCTTCCTTGGAAGTGACGGTGTTCGTGATGTCGAAAACAACGAG 1546
              ||||| |||||||||| ||| | ||   || |||||||||   ||    || |||
Db       1714 CCTGGAATGATTCAGGTGTTCTTGGGTCATAGTGGAGTTCGTGATACGGATGGTAATGAG 1773

Qy       1547 TTGCCTCGATTAGTTTACGTTTCTCGTGAGAAGAGACCCGGATTTGATCACCATAAGAAG 1606
              || || ||  |||| || ||||||||||||||| | || ||||||||||||||| |||||
```

```
b      1774 TTACCACGTCTAGTGTATGTTTCTCGTGAGAAGCGGCCTGGATTTGATCACCACAAGAAA 1833
y      1607 GCTGGAGCTATGAATTCCCTGATACGAGTCTCTGGGGTTCTATCAAATGCTCCTTACCTT 1666
b      1834 GCTGGAGCTATGAATTCCTTGATCCGAGTCTCTGCTGTTCTATCAAACGCTCCTTACCTT 1893
y      1667 CTGAATGTCGATTGTGATCACTACATCAACAATAGCAAAGCTCTTAGAGAAGCAATGTGT 1726
b      1894 CTTAATGTCGATTGTGATCACTACATCAACAACAGCAAAGCAATTAGAGAATCTATGTGT 1953
y      1727 TTCATGATGGATCCTCAGTCAGGAAAGAAAATCTGTTATGTTCAGTTCCCTCAAAGGTTC 1786
b      1954 TTCATGATGGACCCGCAATCGGGAAAGAAAGTTTGTTATGTTCAGTTTCCGCAGAGATTT 2013
y      1787 GATGGGATTGATAGGCACGATCGATACTCAAATCGCAATGTTGTGTTCTTTGATATCAAT 1846
b      2014 GATGGGATTGATAGACATGATAGATACTCAAACCGTAACGTTGTGTTCTTTGATATTAAC 2073
y      1847 ATGAAAGGTTTGGATGGGCTACAAGGGCCTATATACGTCGGTACAGGTTGTGTTTTCAGG 1906
b      2074 ATGAAAGGTCTTGATGGGATACAAGGACCGATATATGTCGGGACAGGTTGTGTGTTTAGA 2133
y      1907 AGGCAAGCGCTTTACGGATTTGATGCACCGAAGAAGAAGAAGGGCCCACGTAAGACATGC 1966
b      2134 AGACAGGCTCTTTATGGTTTTGATGCACCAAAGAAGAAGAAACCACCAGGCAAAACCTGT 2193
y      1967 AATTGCTGGCCAAAATGGTGTCTCCTATGTTTTGGTTCAAGAAAGAATCGTAAAGCAAAG 2026
b      2194 AACTGTTGGCCTAAATGGTGTTGTTTGTGTTGTGGGTTGAGAAAGAAGAGTAAAACGAAA 2253
y      2027 ACAGTGGCTGCGGATAAGAAGAAGAAGAATAGGGAAGCGTCAAAGCAGATCCACGCATTA 2086
b      2254 ------GCCAAAGATAAGAAAACTAACACTAAAGAGACTTCAAAGCAGATTCATGCGCTA 2307
y      2087 GAAAATATCGAAGAGGGCCGCGGTCATAAAGTTCTTAACGTAGAACAGTCAACCGAGGCA 2146
b      2308 GAGAATGTCGACGAAGGTGTTATCGTCCCAGTGTCAAATGTTGAGAAGAGATCTGAAGCA 2367
y      2147 ATGCAAATGAAGTTGCAGAAGAAATATGGGCAGTCTCCTGTATTTGTTGCATCTGCGCGT 2206
b      2368 ACACAATTGAAATTGGAGAAGAAGTTTGGACAATCTCCGGTTTTCGTTGCCTCTGCTGTT 2427
y      2207 CTGGAGAATGGTGGGATGGCTAGAAACGCAAGCCCGGCTTGTCTGCTTAAAGAAGCCATC 2266
b      2428 CTACAGAACGGTGGAGTTCCCCGTAACGCAAGCCCCGCATGTTTGTTAAGAGAAGCCATT 2487
y      2267 CAAGTCATTAGTCGCGGATATGAAGATAAAACTGAATGGGGAAAAGAGATTGGGTGGATC 2326
b      2488 CAAGTTATTAGCTGCGGGTACGAAGATAAAACCGAATGGGGAAAAGAGATCGGGTGGATT 2547
y      2327 TATGGTTCTGTTACCGAAGATATTCTTACGGGTTCTAAGATGCATTCTCATGGTTGGAGA 2386
b      2548 TATGGATCGGTGACTGAAGATATCCTGACGGGTTTCAAGATGCATTGCCATGGATGGAGA 2607
y      2387 CATGTTTATTGTACACCAAAGTTAGCGGCTTTCAAAGGATCAGCTCCAATCAATCTTTCG 2446
b      2608 TCTGTGTACTGTATGCCTAAGCGTGCAGCTTTTAAAGGATCTGCTCCTATTAACTTGTCA 2667
y      2447 GATCGTCTCCATCAAGTTCTTCGATGGGCGCTTGGGTCGGTTGAGATTTTCTTGAGTAGG 2506
b      2668 GATCGTCTTCATCAAGTTCTACGTTGGGCTCTTGGCTCTGTAGAGATTTTCTTGAGCAGA 2727
y      2507 CATTGTCCTATTTGGTATGGTTATGGAGGTGGGTTGAAATGGCTTGAGCGGTTGTCCTAC 2566
b      2728 CATTGTCCGATATGGTATGGTTATGGTGGTGGTTTAAAATGGTTGGAGAGATTCTCTTAC 2787
y      2567 ATTAACTCTGTGGTTTACCCGTGGACCTCTCTACCGCTCATCGTTTACTGTTCTCTCCCT 2626
b      2788 ATCAACTCTGTCGTCTATCCTTGGACTTCACTTCCATTGATCGTCTATTGTTCTCTCCCC 2847
y      2627 GCCATCTGTCTTCTCACTGGAAAATTCATCGTTCCCGAGATTAGCAACTATGCGAGTATC 2686
b      2848 GCGGTTTGTTTACTCACAGGAAAATTCATCGTCCCTGAGATAAGCAACTACGCAGGTATA 2907
```

```
Qy     2687 CTCTTCATGGCGC[illegible] TTCTCGTCGATTGCAATAACGGGTATTCTCGAGATGCAATGGGGC 2746
Db     2908 CTCTTCATGCTCAT[illegible]TTCATATCCATAGCAGTAACTGGAATCCTCGAAATGCAATGGGGA 2967

Qy     2747 AAAGTTGGGATCGA[illegible]GATTGGTGGAGAAACGAACAGTTTTGGGTCATTGGAGGTGTTTCT 2806
Db     2968 GGTGTCGGAATCGA[illegible]ATTGGTGGAGAAACGAACAGTTTTGGGTAATCGGAGGGGCCTCC 3027

Qy     2807 GCGCATCTGTTTG[illegible]TCTTCCAAGGTCTGCTCAAGGTTCTTGCTGGTGTCGACACTAAC 2866
Db     3028 TCGCATCTATTTG[illegible]TGTTTCAAGGTTTGCTCAAAGTTCTAGCCGGAGTTAACACGAAT 3087

Qy     2867 TTCACAGTCACAT[illegible]AAAGCAGCTGATGATGGAGAGTTCTCTGACCTTTACCTCTTCAAA 2926
Db     3088 TTCACAGTCACTT[illegible]AAAGCAGCAGACGATGGAGCTTTCTCTGAGCTTTACATCTTCAAG 3147

Qy     2927 TGGACTTCACTTC[illegible]ATCTCTTCAATGACTC[illegible]ACTCATCATAAACGTCATTGGAGTCATA 2986
Db     3148 TGGACAACTTTGT[illegible]ATTCCTCCGAGAACAC[illegible]TGATCATTAACATCATTGGAGTTATT 3207

Qy     2987 GTCGGAGTCTCTGA[illegible]GCCATCAGCAATGGATACGACTCGTGGGGACCGCTTTTCGGAAGA 3046
Db     3208 GTCGGCGTTTCTGA[illegible]GCCATTAGCAATGGCTATGACTCATGGGGACCTCTCTTTGGGAGA 3267

Qy     3047 CTGTTCTTTGCACTTTGGGTCATCATTCATCTTTACCCGTTCCTTAAAGGTTTGCTTGGG 3106
Db     3268 CTTTTCTTCGCTCTTTGGGTCATTGTTCATTTATACCCATTCCTCAAGGGAATGCTTGGG 3327

Qy     3107 AAACAAGATAGAATG[illegible]CCAACCATTATTGTCGTCTGGTCCATCCTCCTGGCCTCGATTCTT 3166
Db     3328 AAGCAAGACAAAAT[illegible]CCTACGATTATTGTGGTCTGGTCTATTCTTCTAGCTTCGATCTTG 3387

Qy     3167 ACACTTCTTTGGGT[illegible]CGGGTTAATCCGTTTGTGGCGAAAGGCGGTCCTATTCTCGAGATC 3226
Db     3388 ACACTCTTGTGGGT[illegible]AGAGTTAACCCGTTTGTGGCTAAAGGGGGACCAGTGTTGGAGATC 3447

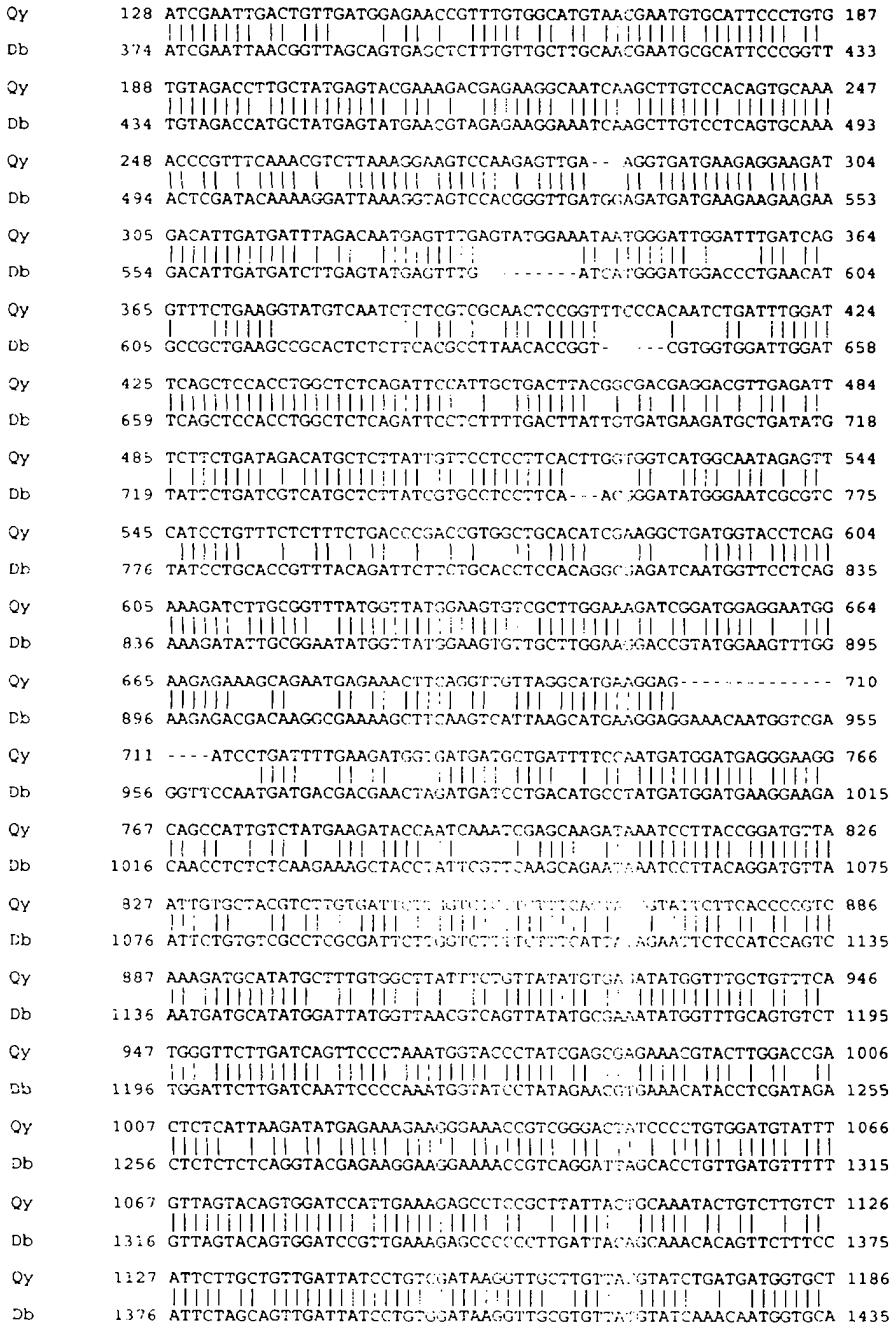
Qy     3227 TGTGGTTTAGACTG 3240
Db     3448 TGTGGTCTGAATTG 3461
```

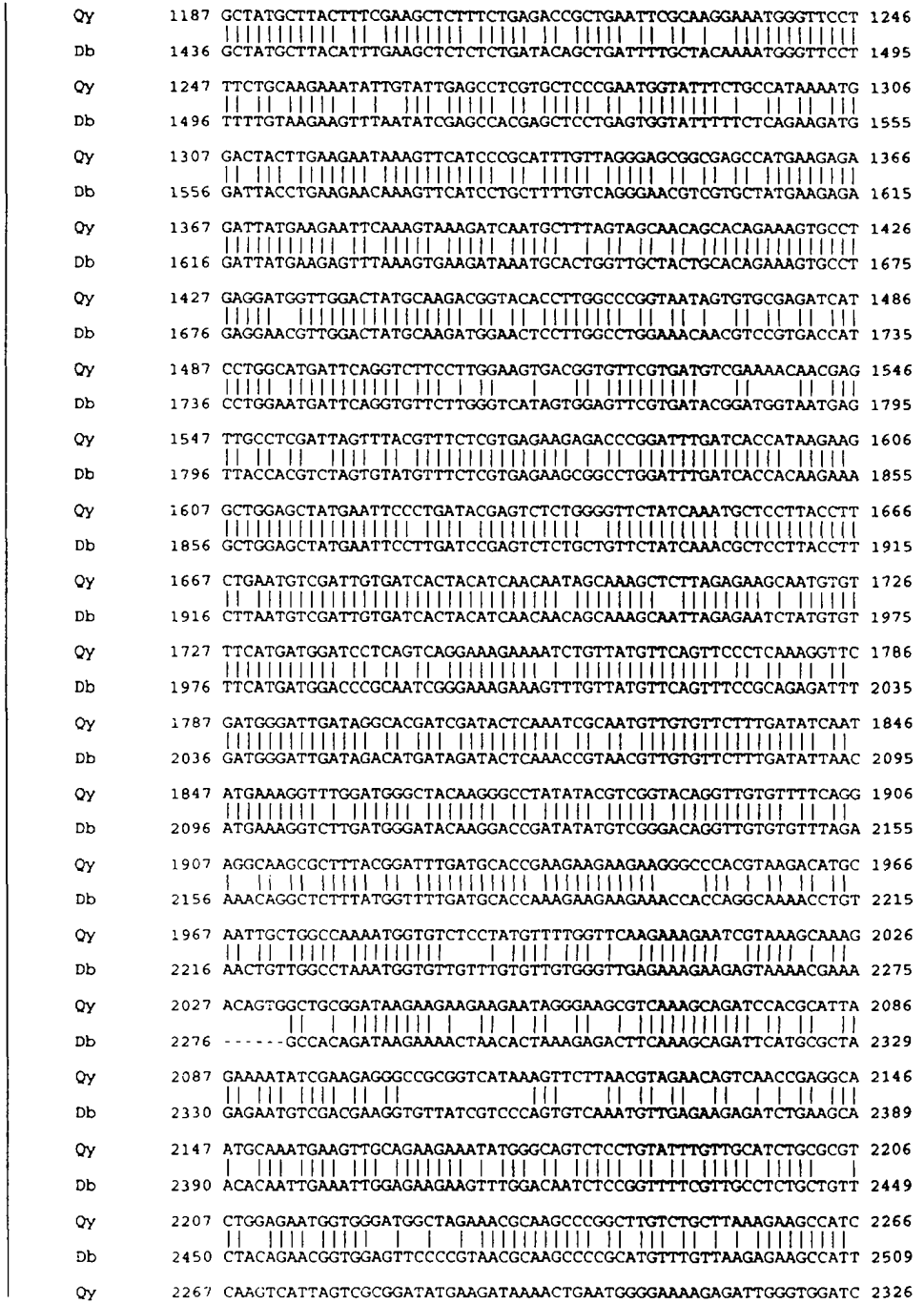
```
RESULT 6
AR267558
LOCUS       AR267558               3828 bp    DNA     linear   PAT 10-APR 2003
DEFINITION  Sequence 7 from patent US 6495740.
ACCESSION   AR267558
VERSION     AR267558.1  GI:29697634
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 3828)
  AUTHORS   Arioli,A., Williamson,R.E., Betzner,A.S. and Peng,L.
  TITLE     Manipulation of cellulose and/or beta-1,4-Glucan
  JOURNAL   Patent: US 6495740-A 7 17-DEC-2002;
FEATURES             Location/Qualifiers
     source          1. .3828
                     /organism="unknown"
BASE COUNT     1056 a    721 c    912 g    1139 t
ORIGIN

  Query Match           56.2%;  Score 1934.4;  DB 6;  Length 3828;
  Best Local Similarity  76.2%;  Pred. No. 0;
  Matches 2478;  Conservative    0;  Mismatches  731;  Indels   45;  Gaps    6;

Qy          8 CGGTTAATCGCCGGTTCTCACAACAGGAATGAGTTTGTCCTCATTAATGCCGATGAGAAT 67
              ||| | || || || ||||||||||| || || || || |||||||| ||||||||| |
Db        254 CGGCTCATTGCTGGCTCTCACAACAGAAACGAATTCGTTCTCATTAACGCCGATGAGAGT 313

Qy         68 GCCCGAATAAGATCAGTCCAAGAGCTGAGTGGACAGACATGTCAAATCTGCAGAGATGAG 127
              ||| ||||| ||||||| ||||| |||||||| || ||||||||||||||  |||||||
Db        314 GCCAGAATACGATCAGTACAAGAACTGAGTGGGCAAACATGTCAAATCTGTGGAGATGAA 373
```





```
b       2510 CAAGTTATTAGCTGCGGGTACCAAGATAAAACCGAATGGGGAAAAGAGATCGGGTGGATT 2569
y       2327 TATGGTTCTGTTACCGAAGATATTCTTACGGGTTCTAAGATGCATTCTCATGGTTGGAGA 2386
b       2570 TATGGATCGGTGACTGAAGATATCCTGACGGGTTTCAAGATGCATTGCCATGGATGGAGA 2629
y       2387 CATGTTTATTGTACACCAAAGTTAGCGGCTTTCAAAGGATCAGCTCCAATCAATCTTTCG 2446
b       2630 TCTGTGTACTGTATGCCTAAGCGTGCAGCTTTTAAAGGATCTGCTCCTATTAACTTGTCA 2689
y       2447 GATCGTCTCCATCAAGTTCTTCGATGGGCGCTTGGGTCGGTTGAGATTTTCTTGAGTAGG 2506
b       2690 GATCGTCTTCATCAAGTTCTACGTTGGGCTCTTGGCTCTGTAGAGATTTTCTTGAGCAGA 2749
y       2507 CATTGTCCTATTTGGTATGGTTATGGAGGTGGGTTGAAATGGCTTGAGCGGTTGTCCTAC 2566
b       2750 CATTGTCCGATATGGTATGGTTATGGTGGTGGTTTAAAATGGTTGGAGAGATTCTCTTAC 2809
y       2567 ATTAACTCTGTGGTTTACCCGTGGACCTCTCTACCGCTCATCGTTTACTGTTCTCTCCCT 2626
b       2810 ATCAACTCTGTCGTCTATCCTTGGACTTCACTTCCATTGATCGTCTATTGTTCTCTCCCC 2869
y       2627 GCCATCTGTCTTCTCACTGGAAAATTCATCGTTCCCGAGATTAGCAACTATGCGAGTATC 2686
b       2870 GCGGTTTGTTTACTCACAGGAAAATTCATCGTCCCTGAGATAAGCAACTACGCAGGTATA 2929
y       2687 CTCTTCATGGCGCTCTTCTCGTCGATTGCAATAACGGGTATTCTCGAGATGCAATGGGGC 2746
b       2930 CTCTTCATGCTCATGTTCATATCCATAGCAGTAACTGGAATCCTCGAAATGCAATGGGGA 2989
y       2747 AAAGTTGGGATCGATGATTGGTGGAGAAACGAACAGTTTTGGGTCATTGGAGGTGTTTCT 2806
b       2990 GGTGTCGGAATCGATGATTGGTGGAGAAACGAGCAGTTTTGGGTAATCGGAGGGGCCTCC 3049
y       2807 GCGCATCTGTTTGCTCTCTTCCAAGGTCTCCTCAAGGTTCTTGCTGGTGTCGACACTAAC 2866
b       3050 TCGCATCTATTTGCTCTGTTTCAAGGTTTGCTCAAAGTTCTAGCCGGAGTTAACACGAAT 3109
y       2867 TTCACAGTCACATCAAAAGCAGCTGATGATGGAGAGTTCTCTGACCTTTACCTCTTCAAA 2926
b       3110 TTCACAGTCACTTCAAAAGCAGCAGACGATGGAGCTTTCTCTGAGCTTTACATCTTCAAG 3169
y       2927 TGGACTTCACTTCTCATCCCTCCAATGACTCTACTCATCATAAACGTCATTGGAGTCATA 2986
b       3170 TGGACAACTTTGTTGATTCCTCCGACAACACTTCTGATCATTAACATCATTGGAGTTATT 3229
y       2987 GTCGGAGTCTCTGATGCCATCAGCAATGGATACGACTCGTGGGGACCGCTTTTCGGAAGA 3046
b       3230 GTCGGCGTTTCTGATGCCATTAGCAATGGCTATGACTCATGGGGACCTCTCTTTGGGAGA 3289
y       3047 CTGTTCTTTGCACTTTGGGTCATCATTCATCTTTACCCGTTCCTTAAAGGTTTGCTTGGG 3106
b       3290 CTTTTCTTCGCTCTTTGGGTCATTGTTCATTTATACCCATTCCTCAAGGGAATGCTTGGG 3349
y       3107 AAACAAGATAGAATGCCAACCATTATTGTCGTCTGGTCCATCCTCCTGGCCTCGATTCTT 3166
b       3350 AAGCAAGACAAAATGCCTACGATTATTGTGGTCTGGTCTATTCTTCTAGCTTCGATCTTG 3409
y       3167 ACACTTCTTTGGGTCCGGGTTAATCCGTTTGTGGCGAAAGGCGGTCCTATTCTCGAGATC 3226
b       3410 ACACTCTTGTGGGTCAGAATTAACCCGTTTGTGGCTAAAGGGGGACCAGTGTTGGAGATC 3469
y       3227 TGTGGTTTAGACTG 3240
b       3470 TGTGGTCTGAATTG 3483

ESULT 7
X030944
OCUS       AX030944               3828 bp    DNA     linear   PAT 24-NOV-2000
EFINITION  Sequence 7 from Patent WO9800549.
CCESSION   AX030944 AX030958
VERSION    AX030944.1  GI:10278347
KEYWORDS   .
SOURCE     Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE  1
  AUTHORS   Williamson,R.E., Peng,L., Arioli,A. and Betzner,A.S.
  TITLE     Manipulation of cellulose and/or beta -1,4-glucan
  JOURNAL   Patent: WO 9800549-A 7 08-JAN-1998;
            WILLIAMSON RICHARD EDWARD (AU) ; PENG LIANGCAI (AU) ; ARIOLI
            ANTONIO (AU) ; UNIV AUSTRALIAN (AU) ; BETZNER ANDREAS STEFAN (AU) ;
            COMMW SCIENT IND RES ORG (AU)
COMMENT    On Oct 15, 2002 this sequence version replaced gi:10278361.
FEATURES             Location/Qualifiers
     source          1. .3828
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /strain="COLUMBIA"
                     /db_xref="taxon:3702"
                     /clone="ATH-A"
     CDS             239. .3493
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAC09626.1"
                     /db_xref="GI:10278348"
                     /db_xref="REMTREMBL:CAC09626"
                     /translation="MNTGGRLIAGSHNRNEFVLINADESARIRSVQELSGQTCQICGD
                     EIELTVSSELFVACNECAFPVCRPCYEYERREGNQACPQCKTRYKRIKGSPRVDGDDE
                     EEEDIDDLEYEFDHGMDPEHAAEAALSSRLNTGRGGLDSAPPGSQIPLLTYCDEDADM
                     YSDRHALIVPPSTGYGNRVYPAPFTDSSAPPQARSMVPQKDIAEYGYGSVAWKDRMEV
                     WKRRQGEKLQVIKHEGGNNGRGSNDDELDDPDMPMMDEGRQPLSRKLPIRSSRINPY
                     RMLILCRLAILGLFFHYRILHPVNDAYGLWLTSVICEIWFAVSWILDQFPKWYPIERE
                     TYLDRLSLRYEKEGKPSGLAPVDVFVSTVDPLKEPPLITANTVLSILAVDYPVDKVAC
                     YVSNNGAAMLTFEALSDTADFATKWVPFCKKFNIEPRAPEWYFSQKMDYLKNKVHPAF
                     VRERRAMKRDYEEFKVKINALVATAQKVPEERWTMQDGTPWPGNNVRDHPGMIQVFLG
                     HSGVRDTDGNELPRLVYVSREKRPGFDHHKKAGAMNSLIRVSAVLSNAPYLLNVDCDH
                     YINNSKAIRESMCFMMDPQSGKKVCYVQFPQRFDGIDRHDRYSNRNVVFFDINMKGLD
                     GIQGPIYVGTGCVFRKQALYGFDAPKKKKPPGKTCNCWPKWCCLCCGLRKKSKTKATD
                     KKTNTKETSKQIHALENVDEGVIVPVSNVEKRSEATQLKLEKKFGQSPVFVASAVLQN
                     GGVPRNASPACLLREAIQVISCGYQDKTEWGKEIGWIYGSVTEDILTGFKMHCHGWRS
                     VYCMPKRAAFKGSAPINLSDRLHQVLRWALGSVEIFLSRHCPIWYGYGGGLKWLERFS
                     YINSVVYPWTSLPLIVYCSLPAVCLLTGKFIVPEISNYAGILFMLMFISIAVTGILEM
                     QWGGVGIDDWWRNEQFWVIGGASSHLFALFQGLLKVLAGVNTNFTVTSKAADDGAFSE
                     LYIFKWTTLLIPPTTLLIINIIGVIVGVSDAISNGYDSWGPLFGRLFFALWVIVHLYP
                     FLKGMLGKQDKMPTIIVVWSILLASILTLLWVRINPFVAKGGPVLEICGLNCGN"
BASE COUNT     1056 a    721 c    912 g   1139 t
ORIGIN

  Query Match           56.2%;  Score 1934.4;  DB 6;  Length 3828;
  Best Local Similarity  76.2%;  Pred. No. 0;
  Matches 2478;  Conservative    0;  Mismatches  731;  Indels   45;  Gaps    6;

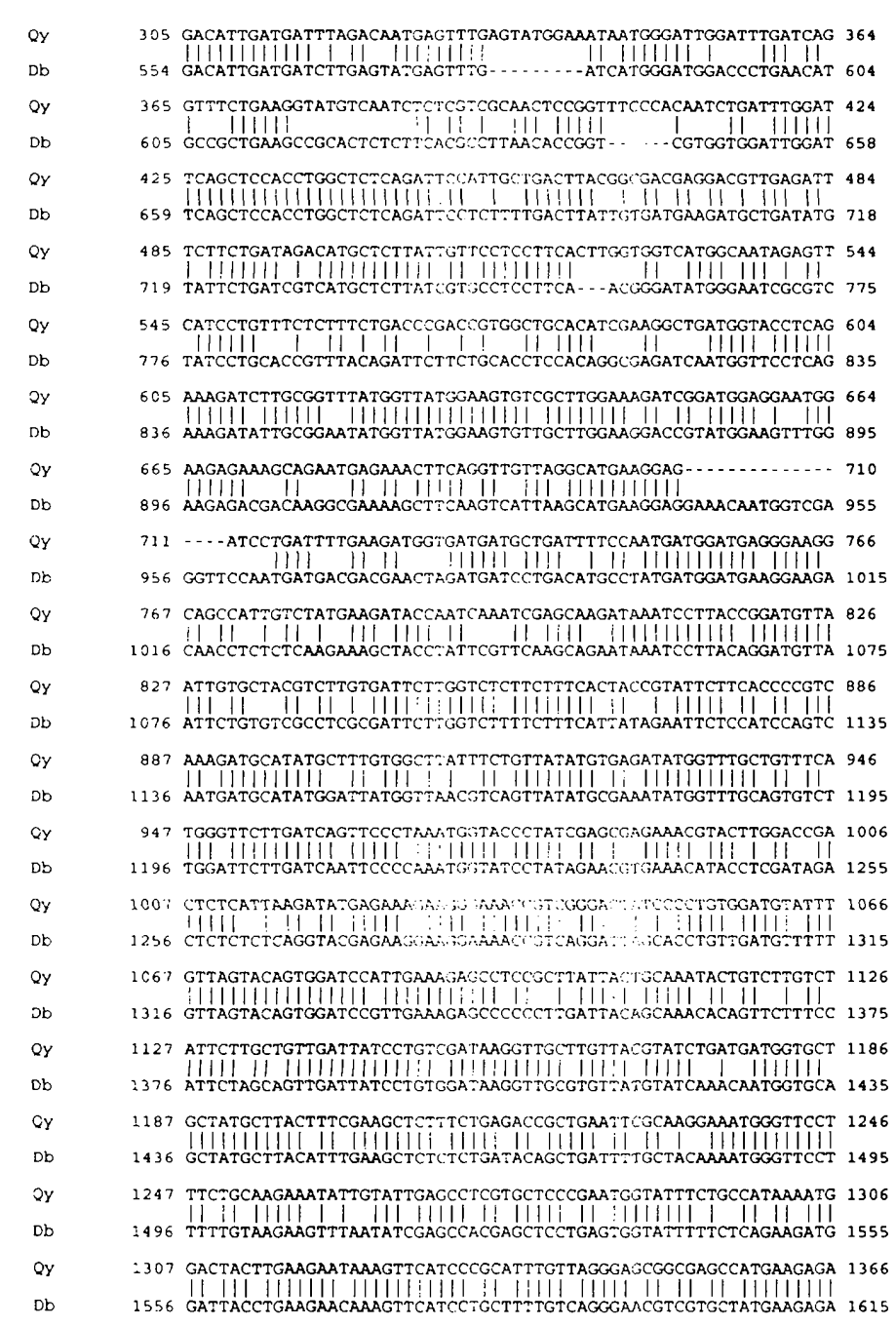
Qy          8 CGGTTAATCGCCGGTTCTCACAACAGGAATGAGTTTGTCCTCATTAATGCCGATGAGAAT 67
Db        254 CGGCTCATTGCTGGCTCTCACAACAGAAACGAATTCGTTCTCATTAACGCCGATGAGAGT 313

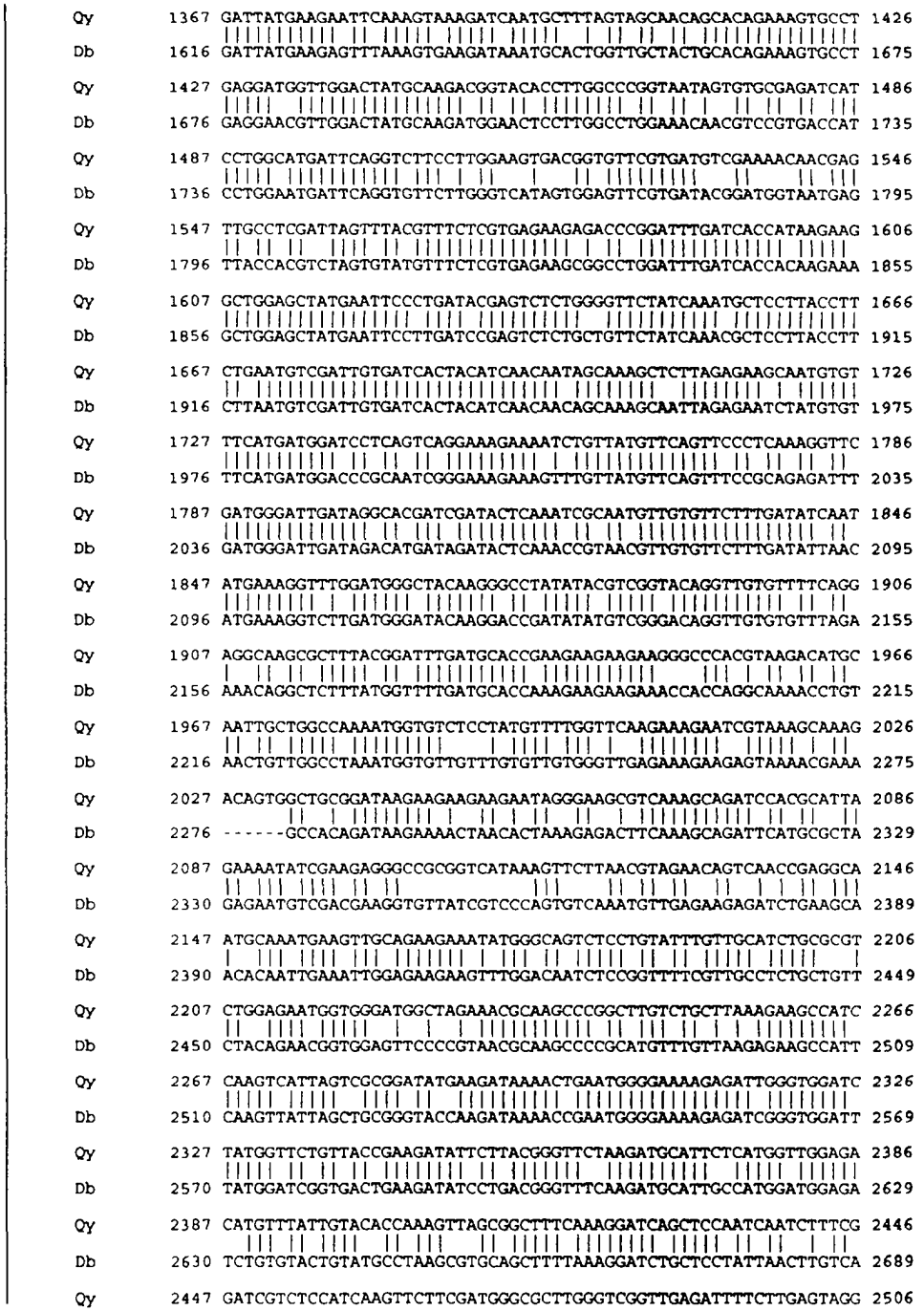
Qy         68 GCCCGAATAAGATCAGTCCAAGAGCTGAGTGGACAGACATGTCAAATCTGCAGAGATGAG 127
Db        314 GCCAGAATACGATCAGTACAAGAACTGAGTGGGCAAACATGTCAAATCTGTGGAGATGAA 373

Qy        128 ATCGAATTGACTGTTGATGGAGAACCGTTTGTGGCATGTAACGAATGTGCATTCCCTGTG 187
Db        374 ATCGAATTAACGGTTAGCAGTGAGCTCTTTGTTGCTTGCAACGAATGCGCATTCCCGGTT 433

Qy        188 TGTAGACCTTGCTATGAGTACGAAAGACGAGAAGGCAATCAAGCTTGTCCACAGTGCAAA 247
Db        434 TGTAGACCATGCTATGAGTATGAACGTAGAGAAGGAAATCAAGCTTGTCCTCAGTGCAAA 493

Qy        248 ACCCGTTTCAAACGTCTTAAAGGAAGTCCAAGAGTTGA---AGGTGATGAAGAGGAAGAT 304
Db        494 ACTCGATACAAAAGGATTAAAGGTAGTCCACGGGTTGATGGAGATGATGAAGAAGAAGAA 553
```





```
b       2690 GATCGTCTTCATCAAGTTCTACGTTGGGCTCTTGGCTCTGTAGAGATTTTCTTGAGCAGA 2749

y       2507 CATTGTCCTATTTGGTATGGTTATGGAGGTGGGTTGAAATGGCTTGAGCGGTTGTCCTAC 2566

b       2750 CATTGTCCGATATGGTATGGTTATGGTGGTGGTTTAAAATGGTTGGAGAGATTCTCTTAC 2809

y       2567 ATTAACTCTGTGGTTTACCCGTGGACCTCTCTACCGCTCATCGTTTACTGTTCTCTCCCT 2626

b       2810 ATCAACTCTGTCGTCTATCCTTGGACTTCACTTCCATTGATCGTCTATTGTTCTCTCCCC 2869

y       2627 GCCATCTGTCTTCTCACTGGAAAATTCATCGTTCCCGAGATTAGCAACTATGCGAGTATC 2686

b       2870 GCGGTTTGTTTACTCACAGGAAAATTCATCGTCCCTGAGATAAGCAACTACGCAGGTATA 2929

y       2687 CTCTTCATGGCGCTCTTCTCGTCGATTGCAATAACGGGTATTCTCGAGATGCAATGGGGC 2746

b       2930 CTCTTCATGCTCATGTTCATATCCATAGCAGTAACTGGAATCCTCGAAATGCAATGGGGA 2989

y       2747 AAAGTTGGGATCGATGATTGGTGGAGAAACGAACAGTTTTGGGTCATTGGAGGTGTTTCT 2806

b       2990 GGTGTCGGAATCGATGATTGGTGGAGAAACGAGCAGTTTTGGGTAATCGGAGGGGCCTCC 3049

y       2807 GCGCATCTGTTTGCTCTCTTCCAAGGTCTCCTCAAGGTTCTTGCTGGTGTCGACACTAAC 2866

b       3050 TCGCATCTATTTGCTCTGTTTCAAGGTTTGCTCAAAGTTCTAGCCGGAGTTAACACGAAT 3109

y       2867 TTCACAGTCACATCAAAAGCAGCTGATGATGGAGAGTTCTCTGACCTTTACCTCTTCAAA 2926

b       3110 TTCACAGTCACTTCAAAAGCAGCAGACGATGGAGCTTTCTCTGAGCTTTACATCTTCAAG 3169

y       2927 TGGACTTCACTTCTCATCCCTCCAATGACTCTACTCATCATAAACGTCATTGGAGTCATA 2986

b       3170 TGGACAACTTTGTTGATTCCTCCGACAACACTTCTGATCATTAACATCATTGGAGTTATT 3229

y       2987 GTCGGAGTCTCTGATGCCATCAGCAATGGATACGACTCGTGGGGACCGCTTTTCGGAAGA 3046

b       3230 GTCGGCGTTTCTGATGCCATTAGCAATGGCTATGACTCATGGGGACCTCTCTTTGGGAGA 3289

y       3047 CTGTTCTTTGCACTTTGGGTCATCATTCATCTTTACCCGTTCCTTAAAGGTTTGCTTGGG 3106

b       3290 CTTTTCTTCGCTCTTTGGGTCATTGTTCATTTATACCCATTCCTCAAGGGAATGCTTGGG 3349

y       3107 AAACAAGATAGAATGCCAACCATTATTGTCGTCTGGTCCATCCTCCTGGCCTCGATTCTT 3166

b       3350 AAGCAAGACAAAATGCCTACGATTATTGTGGTCTGGTCTATTCTTCTAGCTTCGATCTTG 3409

y       3167 ACACTTCTTTGGGTCCGGGTTAATCCGTTTGTGGCGAAAGGCGGTCCTATTCTCGAGATC 3226

b       3410 ACACTCTTGTGGGTCAGAATTAACCCGTTTGTGGCTAAAGGGGGACCAGTGTTGGAGATC 3469

y       3227 TGTGGTTTAGACTG 3240

b       3470 TGTGGTCTGAATTG 3483
```

```
ESULT 8
D022677
OCUS        BD022677               3828 bp    DNA     linear   PAT 27-AUG-2002
EFINITION   Manipulation of cellulose and/or beta-1,4-glucan.
CCESSION    BD022677
ERSION      BD022677.1  GI:22563900
EYWORDS     JP 2001510326-A/5.
OURCE       Arabidopsis thaliana (thale cress)
 ORGANISM   Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
EFERENCE    1  (bases 1 to 3828)
 AUTHORS    Ariori,A., Williamson,R.E., Bettsunar,A.S. and Penn,L.
 TITLE      Manipulation of cellulose and/or beta-1,4-glucan
 JOURNAL    Patent: JP 2001510326-A 5 31-JUL-2001;
            THE AUSTRALIAN NATIONAL UNIVERSITY, COMMONWEALTH SCIENTIFIC AND
            INDUSTRIAL RESEARCH ORGANIZATION
COMMENT     OS   Arabidopsis thaliana (thale cress)
            PN   JP 2001510326 A/5
            PD   31-JUL-2001
            PF   24-JUN-1997 JP 1998503665
            PR   27-JUN-1996 AU      PO 0699
            PI   ANTONIO ARIORI,RICHARD EDWARD WILLIAMSON,ANDREAS STEPHAN  PI
            BETTSUNAR,
            PI   LYONKAI PENN
            PC   C12N15/09,A01H5/00,C07K16/40,C12N1/15,C12N1/19,C12N1/21,C12N5/
            PC   10,C12N5/10,
            PC   C12N9/10,C12N15/00,C12N5/00,C12N5/00
            CC   Strandedness: Single;
            CC   Topology: Linear;
            FH   Key             Location/Qualifiers
            FT   CDS             239..3498
FEATURES            Location/Qualifiers
     source         1..3828
                    /organism="Arabidopsis thaliana"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:3702"
BASE COUNT     1056 a     721 c     912 g    1139 t
ORIGIN
```

Query Match 56.2%; Score 1934.4; DB 6; Length 3828;
Best Local Similarity 76.2%; Pred. No. 0;
Matches 2478; Conservative 0; Mismatches 731; Indels 45; Gaps 6;

```
Qy          8 CGGTTAATCGCCGGTTCTCACAACAGGAATGAGTTTGTCCTCATTAATGCCGATGAGAAT 67

Db        254 CGGCTCATTGCTGGCTCTCACAACAGAAACGAATTCGTTCTCATTAACGCCGATGAGAGT 313

Qy         68 GCCCGAATAAGATCAGTCCAAGAGCTGAGTGGACAGACATGTCAAATCTGCAGAGATGAG 127

Db        314 GCCAGAATACGATCAGTACAAGAACTGAGTGGGCAAACATGTCAAATCTGTGGAGATGAA 373

Qy        128 ATCGAATTGACTGTTGATGGAGAACCGTTTGTGGCATGTAACGAATGTGCATTCCCTGTG 187

Db        374 ATCGAATTAACGGTTAGCAGTGAGCTCTTTGTTGCTTGCAACGAATGCGCATTCCCGGTT 433

Qy        188 TGTAGACCTTGCTATGAGTACGAAAGACGAGAAGGCAATCAAGCTTGTCCACAGTGCAAA 247

Db        434 TGTAGACCATGCTATGAGTATGAACGTAGAGAAGGAAATCAAGCTTGTCCTCAGTGCAAA 493

Qy        248 ACCCGTTTCAAACGTCTTAAAGGAAGTCCAAGAGTTGA---AGGTGATGAAGAGGAAGAT 304

Db        494 ACTCGATACAAAAGGATTAAAGGTAGTCCACGGGTTGATGGAGATGATGAAGAAGAAGAA 553

Qy        305 GACATTGATGATTTAGACAATGAGTTTGAGTATGGAAATAATGGGATTGGATTTGATCAG 364

Db        554 GACATTGATGATCTTGAGTATGAGTTTG- -------ATCATGGGATGGACCCTGAACAT 604

Qy        365 GTTTCTGAAGGTATGTCAATCTCTCGTCGCAACTCCGGTTTCCCACAATCTGATTTGGAT 424

Db        605 GCCGCTGAAGCCGCACTCTCTTCACGCCTTAACACCGGT----- CGTGGTGGATTGGAT 658

Qy        425 TCAGCTCCACCTGGCTCTCAGATTCCATTGCTGACTTACGGCGACGAGGACGTTGAGATT 484

Db        659 TCAGCTCCACCTGGCTCTCAGATTCCTCTTTTGACTTATTGTGATGAAGATGCTGATATG 718

Qy        485 TCTTCTGATAGACATGCTCTTATTGTTCCTCCTTCACTTGGTGGTCATGGCAATAGAGTT 544

Db        719 TATTCTGATCGTCATGCTCTTATCGTGCCTCCTTCA---ACGGGATATGGGAATCGCGTC 775

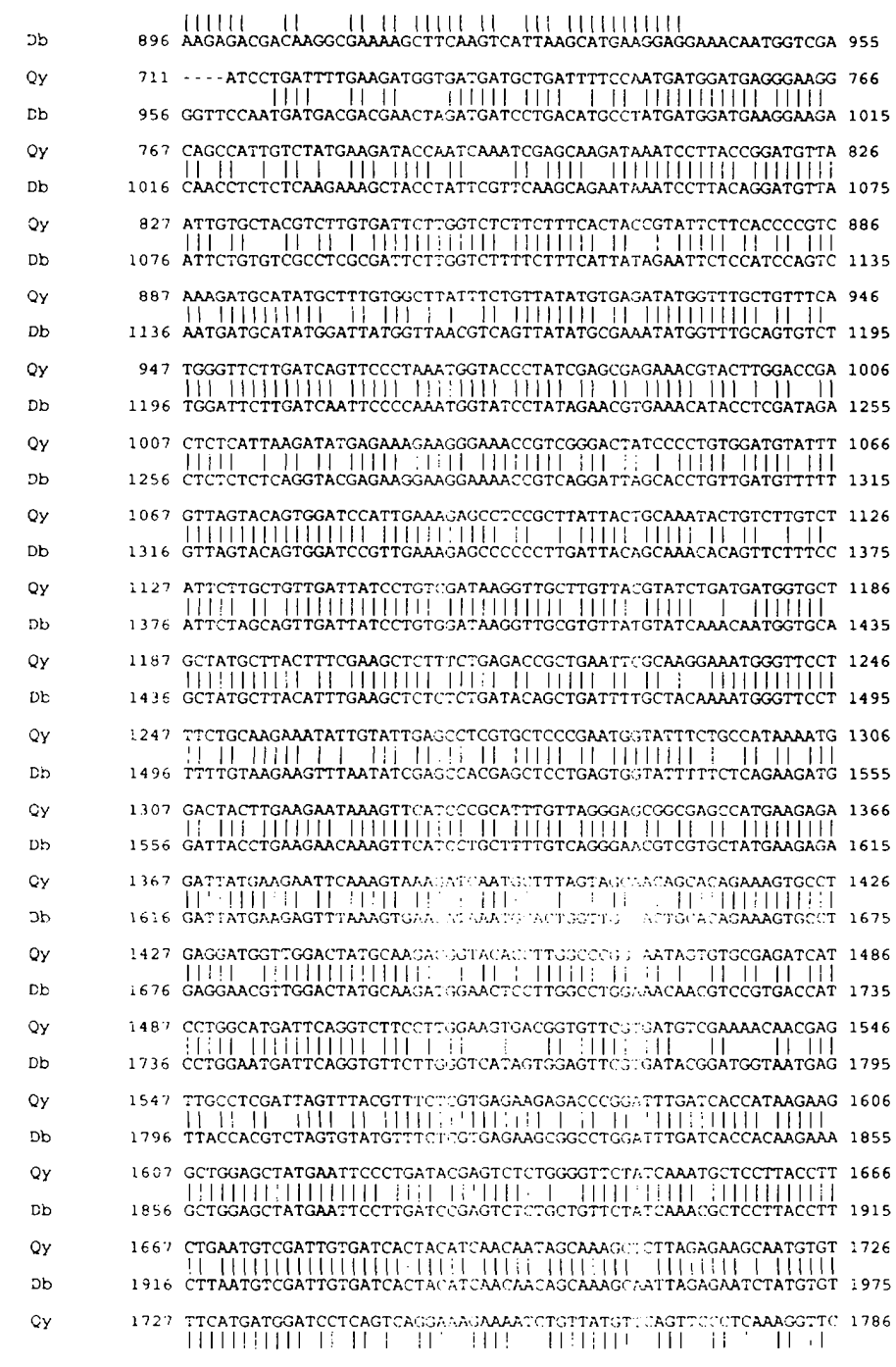
Qy        545 CATCCTGTTTCTCTTTCTGACCCGACCGTGGCTGCACATCGAAGGCTGATGGTACCTCAG 604

Db        776 TATCCTGCACCGTTTACAGATTCTTCTGCACCTCCACAGGCGAGATCAATGGTTCCTCAG 835

Qy        605 AAAGATCTTGCGGTTTATGGTTATGGAAGTGTCGCTTGGAAAGATCGGATGGAGGAATGG 664

Db        836 AAAGATATTGCGGAATATGGTTATGGAAGTGTTGCTTGGAAGGACCGTATGGAAGTTTGG 895

Qy        665 AAGAGAAAGCAGAATGAGAAACTTCAGGTTGTTAGGCATGAAGGAG---------- -- 710
```


Db 896 AAGAGACGACAAGGCGAAAAGCTTCAAGTCATTAAGCATGAAGGAGGAAACAATGGTCGA 955
Qy 711 ----ATCCTGATTTTGAAGATGGTGATGATGCTGATTTTCCAATGATGGATGAGGGAAGG 766
Db 956 GGTTCCAATGATGACGACGAACTAGATGATCCTGACATGCCTATGATGGATGAAGGAAGA 1015
Qy 767 CAGCCATTGTCTATGAAGATACCAATCAAATCGAGCAAGATAAATCCTTACCGGATGTTA 826
Db 1016 CAACCTCTCTCAAGAAAGCTACCTATTCGTTCAAGCAGAATAAATCCTTACAGGATGTTA 1075
Qy 827 ATTGTGCTACGTCTTGTGATTCTTGGTCTCTTCTTTCACTACCGTATTCTTCACCCCGTC 886
Db 1076 ATTCTGTGTCGCCTCGCGATTCTTGGTCTTTTCTTTCATTATAGAATTCTCCATCCAGTC 1135
Qy 887 AAAGATGCATATGCTTTGTGGCTTATTTCTGTTATATGTGAGATATGGTTTGCTGTTTCA 946
Db 1136 AATGATGCATATGGATTATGGTTAACGTCAGTTATATGCGAAATATGGTTTGCAGTGTCT 1195
Qy 947 TGGGTTCTTGATCAGTTCCCTAAATGGTACCCTATCGAGCGAGAAACGTACTTGGACCGA 1006
Db 1196 TGGATTCTTGATCAATTCCCCAAATGGTATCCTATAGAACGTGAAACATACCTCGATAGA 1255
Qy 1007 CTCTCATTAAGATATGAGAAAGAAGGGAAACCGTCGGGACTATCCCCTGTGGATGTATTT 1066
Db 1256 CTCTCTCTCAGGTACGAGAAGGAAGGAAAACCGTCAGGATTAGCACCTGTTGATGTTTTT 1315
Qy 1067 GTTAGTACAGTGGATCCATTGAAAGAGCCTCCGCTTATTACTGCAAATACTGTCTTGTCT 1126
Db 1316 GTTAGTACAGTGGATCCGTTGAAAGAGCCCCCCTTGATTACAGCAAACACAGTTCTTTCC 1375
Qy 1127 ATTCTTGCTGTTGATTATCCTGTGGATAAGGTTGCTTGTTACGTATCTGATGATGGTGCT 1186
Db 1376 ATTCTAGCAGTTGATTATCCTGTGGATAAGGTTGCGTGTTATGTATCAAACAATGGTGCA 1435
Qy 1187 GCTATGCTTACTTTCGAAGCTCTTTCTGAGACCGCTGAATTCGCAAGGAAATGGGTTCCT 1246
Db 1436 GCTATGCTTACATTTGAAGCTCTCTCTGATACAGCTGATTTTGCTACAAAATGGGTTCCT 1495
Qy 1247 TTCTGCAAGAAATATTGTATTGAGCCTCGTGCTCCCGAATGGTATTTCTGCCATAAAATG 1306
Db 1496 TTTTGTAAGAAGTTTAATATCGAGCCACGAGCTCCTGAGTGGTATTTTTCTCAGAAGATG 1555
Qy 1307 GACTACTTGAAGAATAAAGTTCATCCCGCATTTGTTAGGGAGCGGCGAGCCATGAAGAGA 1366
Db 1556 GATTACCTGAAGAACAAAGTTCATCCTGCTTTTGTCAGGGAACGTCGTGCTATGAAGAGA 1615
Qy 1367 GATTATGAAGAATTCAAAGTAAA
Db 1616 GATTATGAAGAGTTTAAAGTGAA
Qy 1427 GAGGATGGTTGGACTATGCAAGA
Db 1676 GAGGAACGTTGGACTATGCAAGA
Qy 1487 CCTGGCATGATTCAGGTCTTCCT
Db 1736 CCTGGAATGATTCAGGTGTTCTT
Qy 1547 TTGCCTCGATTAGTTTACGTTTC
Db 1796 TTACCACGTCTAGTGTATGTTTC
Qy 1607 GCTGGAGCTATGAATTCCCTGA
Db 1856 GCTGGAGCTATGAATTCCTTGA
Qy 1667 CTGAATGTCGATTGTGATCACT
Db 1916 CTTAATGTCGATTGTGATCACT
Qy 1727 TTCATGATGGATCCTCAGTCAG
Db 1976 TTCATGATGGACCCGCAATCGG
Qy 1787 GATGGGATTGATAGGCACGATC
Db 2036 GATGGGATTGATAGACATGATA
Qy 1847 ATGAAAGGTTTGGATGGGCTAC
Db 2096 ATGAAAGGTCTTGATGGGATAC
Qy 1907 AGGCAAGCGCTTTACGGATTTG
Db 2156 AAACAGGCTCTTTATGGTTTTG
Qy 1967 AATTGCTGGCCAAAATGGTGTC
Db 2216 AACTGTTGGCCTAAATGGTGTT
Qy 2027 ACAGTGGCTGCGGATAAGAAGA
Db 2276 ------GCCACAGATAAGAAAA
Qy 2087 GAAAATATCGAAGAGGGCCGCG
Db 2330 GAGAATGTCGACGAAGGTGTTA
Qy 2147 ATGCAAATGAAGTTGCAGAAGA
Db 2390 ACACAATTGAAATTGGAGAAGA
Qy 2207 CTGGAGAATGGTGGGATGGCTA
Db 2450 CTACAGAACGGTGGAGTTCCCC
Qy 2267 CAAGTCATTAGTCGCGGATATG
Db 2510 CAAGTTATTAGCTGCGGGTACC
Qy 2327 TATGGTTCTGTTACCGAAGATA
Db 2570 TATGGATCGGTGACTGAAGATA
Qy 2387 CATGTTTATTGTACACCAAAGT
Db 2630 TCTGTGTACTGTATGCCTAAGC
Qy 2447 GATCGTCTCCATCAAGTTCTTC
Db 2690 GATCGTCTTCATCAAGTTCTAC
Qy 2507 CATTGTCCTATTTGGTATGGTT
Db 2750 CATTGTCCGATATGGTATGGTT
Qy 2567 ATTAACTCTGTGGTTTACCCGT
Db 2810 ATCAACTCTGTCGTCTATCCTT
Qy 2627 GCCATCTGTCTTCTCACTGGAA
Db 2870 GCGGTTTGTTTACTCACAGGAA
Qy 2687 CTCTTCATGGCGCTCTTCTCGT
Db 2930 CTCTTCATGCTCATGTTCATAT
Qy 2747 AAAGTTGGGATCGATGATTGGT
Db 2990 GGTGTCGGAATCGATGATTGGT
Qy 2807 GCGCATCTGTTTGCTCTCTTCC
Db 3050 TCGCATCTATTTGCTCTGTTTC

```
y       2867 TTCACAGTCACATCAAAAGCAGCTGATGATGGAGAGTTCTCTGACCTTTACCTCTTCAAA 2926
             |||||||||||| |||||||||| || |||||||  |||||||| ||||||| |||||||
b       3110 TTCACAGTCACTTCAAAAGCAGCAGACGATGGAGCTTTCTCTGAGCTTTACATCTTCAAG 3169

y       2927 TGGACTTCACTTCTCATCCCTCCAATGACTCTACTCATCATAAACGTCATTGGAGTCATA 2986
             |||||  |   |  | || ||||| |  || || || ||||| ||| ||||||||||  ||
b       3170 TGGACAACTTTGTTGATTCCTCCGACAACACTTCTGATCATTAACATCATTGGAGTTATT 3229

y       2987 GTCGGAGTCTCTGATGCCATCAGCAATGGATACGACTCGTGGGGACCGCTTTTCGGAAGA 3046
             ||||| || |||||||||| |||||||| || ||||| |||||||| || || || |||
b       3230 GTCGGCGTTTCTGATGCCATTAGCAATGGCTATGACTCATGGGGACCTCTCTTTGGGAGA 3289

y       3047 CTGTTCTTTGCACTTTGGGTCATCATTCATCTTTACCCGTTCCTTAAAGGTTTGCTTGGG 3106
             || ||||| || |||||||||| ||||| | ||||| ||||| || ||  ||||||||
b       3290 CTTTTCTTCGCTCTTTGGGTCATTGTTCATTTATACCCATTCCTCAAGGGAATGCTTGGG 3349

y       3107 AAACAAGATAGAATGCCAACCATTATTGTCGTCTGGTCCATCCTCCTGGCCTCGATTCTT 3166
             || ||||| | |||||| || |||||||| |||||||| || || || || ||||| |
b       3350 AAGCAAGACAAAATGCCTACGATTATTGTGGTCTGGTCTATTCTTCTAGCTTCGATCTTG 3409

y       3167 ACACTTCTTTGGGTCCGGGTTAATCCGTTTGTGGCGAAAGGCGGTCCTATTCTCGAGATC 3226
             |||||  | |||||| |  |||| |||||||||| ||||| || ||  |  | ||||||
b       3410 ACACTCTTGTGGGTCAGAATTAACCCGTTTGTGGCTAAAGGGGGACCAGTGTTGGAGATC 3469

y       3227 TGTGGTTTAGACTG 3240
             |||||| |  | ||
b       3470 TGTGGTCTGAATTG 3483
```

```
ESULT 9
F304374
OCUS       AF304374               3847 bp    mRNA    linear   PLN 02-MAY-2001
EFINITION  Nicotiana alata cellulose synthase catalytic subunit (CesA1) mRNA,
           complete cds.
CCESSION   AF304374
ERSION     AF304374.1  GI:13925880
EYWORDS    .
OURCE      Nicotiana alata (Persian tobacco)
 ORGANISM  Nicotiana alata
           Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
           Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
           Asteridae; lamiids; Solanales; Solanaceae; Nicotiana.
EFERENCE   1  (bases 1 to 3847)
 AUTHORS   Doblin,M.S., De Melis,L., Newbigin,E., Bacic,A. and Read,S.M.
 TITLE     Pollen tubes of Nicotiana alata express two genes from different
           beta-glucan synthase families
 JOURNAL   Plant Physiol. 125 (4), 2040-2052 (2001)
 MEDLINE   21196092
  PUBMED   11299383
EFERENCE   2  (bases 1 to 3847)
 AUTHORS   Doblin,M.S., De Melis,L., Newbigin,E., Bacic,A. and Read,S.M.
 TITLE     Direct Submission
 JOURNAL   Submitted (12-SEP-2000) Plant Biology, University of California, 1
           Shields Ave, Davis, CA 95616, USA
EATURES             Location/Qualifiers
     source         1. .3847
                    /organism="Nicotiana alata"
                    /mol_type="mRNA"
                    /db_xref="taxon:4087"
                    /tissue_type="pistil"
                    /note="genotype: S2S2"
     gene           1. .3847
                    /gene="CesA1"
     CDS            103. .3378
                    /gene="CesA1"
                    /codon_start=1
                    /product="cellulose synthase catalytic subunit"
                    /protein_id="AAK49454.1"
                    /db_xref="GI:13925881"
                    /translation="MDTKGRLVAGSHNRNEFVVINADDVGRVTSVKELSGQICQICGD
                    EIEVTVDGEPFIACNECAFPVCRQCYEYERREGNQACPQCKTRFKRIKGSPRVDGDDE
                    DDEFDDLDHEFDYHGNPRYMSEAAFSSRLGRGTNHNASGLTTPSEVDPAALNSEIPLL
```

```
                    TYGQED[illegible]SADKHALIIPPFMGRGKKVHPVPYSDSMSLPPRPMDPKKDLAVYGYGTV
                    AWKERMK[illegible]WKKKQNDKLQVVKHGGGKGGGNDGDELDDPDLPKMDEGRQPLSRKLPISS
                    SRLSPYRLLILVRLAVVGLFFHYRITHPVNDAYALWLISIICEIWFAVSWIFDQFPKW
                    FPIVRETYLDRLSLRYEKEGKPSGLAPIDIFVSTVDPLKEPPLITANTVLSILAVDYP
                    EDKVSCYVSDDGAAMLTFEALSETSEFARKWVPFCKKFNIEPRAPEWYFSQKVDYLKN
                    KVHPSFVRERRAMKRDYEEFKVRINGLVATAQKVPEDGWTMQDGTPWPGNLVRDHPGM
                    IQVFLGNEGVRDIEGNVLPRLIYVSREKRPGFDHHKKAGAMNALMRVSAVISNAPYLL
                    NVDCDHYINNSKALREAMCFMMDPTSGKKICYVQFPQRFDGIDRHDRYSNRNVVFFDI
                    NMKGLDGIQGPIYVGTGCVFRRQALYGYDAPKKTKPPGKTCNCWPKWCCCCFGSRKKH
                    KKGKTTKDNKKKTKTKEASPQIHALENIEEGIEGIDSEKATLMPQIKLEKKFGQSPVF
                    VASTLLEDGGIPPGATSASLLKEAIHVISCGYEDKTEWGREVGWIYGSVTEDILTGFK
                    MHCHGWRSVYCMPKRPAFKGSAPINLSDRLHQVLRWALGSVEILLSKHCPIWYGYGCG
                    LKPLERFSYINSVVYPLTSLPLIAYCALPAVCLLTGKFIVPEISNYASILFMGLFIMI
                    AATSVLEMQWGGVTIDDWWRNEQFWVIGGASSHLFALFQGLLKVLAGVSTSFTVTSKA
                    ADDGEFSELYLFKWTSLLIPPMTLLIINIIGVIVGISDAINNGYDSWGPLFGRLFFAL
                    WVIVHLYPFLKGVMGRQNKVPTIIVVWSILLASIFSLLWVRVNPFTARGGLVLEVCGL
                    DCE"
BASE COUNT     1061 a    7[illegible]  925 t  [illegible]
ORIGIN

  Query Match          5[illegible]%;  Score 1784;  [illegible] 8;  Length 3847,
  Best Local Similarity  7[illegible]%;  Pred. No. 0;
  Matches 2376;  Conservative    0;  Mismatches  855;  Indels   45;  Gaps    6;

Qy          7 GCGGTTAATCGCCG[illegible]TTCTCACAACAGGAATGAGTTTGTCCTCATTAATGCCGATGAGAA 66
              | || || | || | ||| ||||| ||||||  |||||  |||| ||||| |||||
Db        117 GAGGCTAGTTGCTG[illegible]TTCACACAATAGGAATGAGTTTGTTGTCATCAATGCTGATGATGT 176

Qy         67 TGCCCGAATAAGATCAGTCCAAGAGCTGAGTGGACAGACATGTCAAATCTGCAGAGATGA 126
              ||   || |||      ||  |||| | ||| | ||||  || || || ||  |||||||
Db        177 TGGAAGAGTAACTT[illegible]TGTGAAAGAATTAAGTGGGCAGATTTGCCAGATTTGTGGAGATGA 236

Qy        127 GATCGAATTGACTG[illegible]TGATGGAGAACCGTTT[illegible]TGGCATGTAACGAATGTGCATTCCCTGT 186
              ||| ||| | ||  .  ||||| ||  | ||| | || || ||||||||||||||||||||||
Db        237 GATTGAAGTTACAG[illegible]AGATGGGGAGCCATTTATTGCTTGCAACGAATGTGCATTCCCTGT 296

Qy        187 GTGTAGACCTTGCTATGAGTACGAAAGACGAGAAGGCAATCAAGCTTGTCCACAGTGCAA 246
              || |||| |||| ||||| |||||  -|||||| ||||||||||| || || ||||||
Db        297 TTGCAGACAATGCTATGAGTATGAAAGGAGAGAAGGAAATCAAGCTTGCCCTCAATGCAA 356

Qy        247 AACCCGTTTCAAACGTCTTAAAGGAAGTCCAAGAGTTGAAGGTGATGAAGAGGAAGATGA 306
              ||| || |||| ||  |||||| || |  || |||| || |||| || || |||||
Db        357 AACTCGGTTCAAGCGCATTAAAGGGAGCCCCAGGGTTGATGGAGATGATGAAGATGATGA 416

Qy        307 CATTGATGATTTAGACAATGAGTTTGAGTATGGAAATAA----------------TGGGAT 351
              |||||||  |||||| |||||||  || ||  |||                 || |
Db        417 ATTTGATGACTTAGACCATGAGTTCGATTACCATGGTAACCCTCGATATATGTCTGAGGC 476

Qy        352 TGGATTTGATCAGGTTTCTGAAGGTATGTCAATCTCTCGTCGCAACTCCGGTTTCCCACA 411
              || |||   |   |  ||   | |  |  |  |                |  | |
Db        477 TGCATTCTCTTCTCGTCTTGGCCGCGGTACCAATCATAATGCTTCTGGGCTTACTACCCC 536

Qy        412 ATCTGATTTGGATTCAGCTCCACCTGGCTCTCAGATTCCATTGCTGACTTACGGCGACGA 471
              ||| ||  | ||| | ||  |  | |  |   ||| ||  |  || |||| ||  | ||
Db        537 ATCAGAAGTAGATCCTGCTGCTCTTAACTCAGAGATCCCTCTTCTTACTTATGGTCAAGA 596

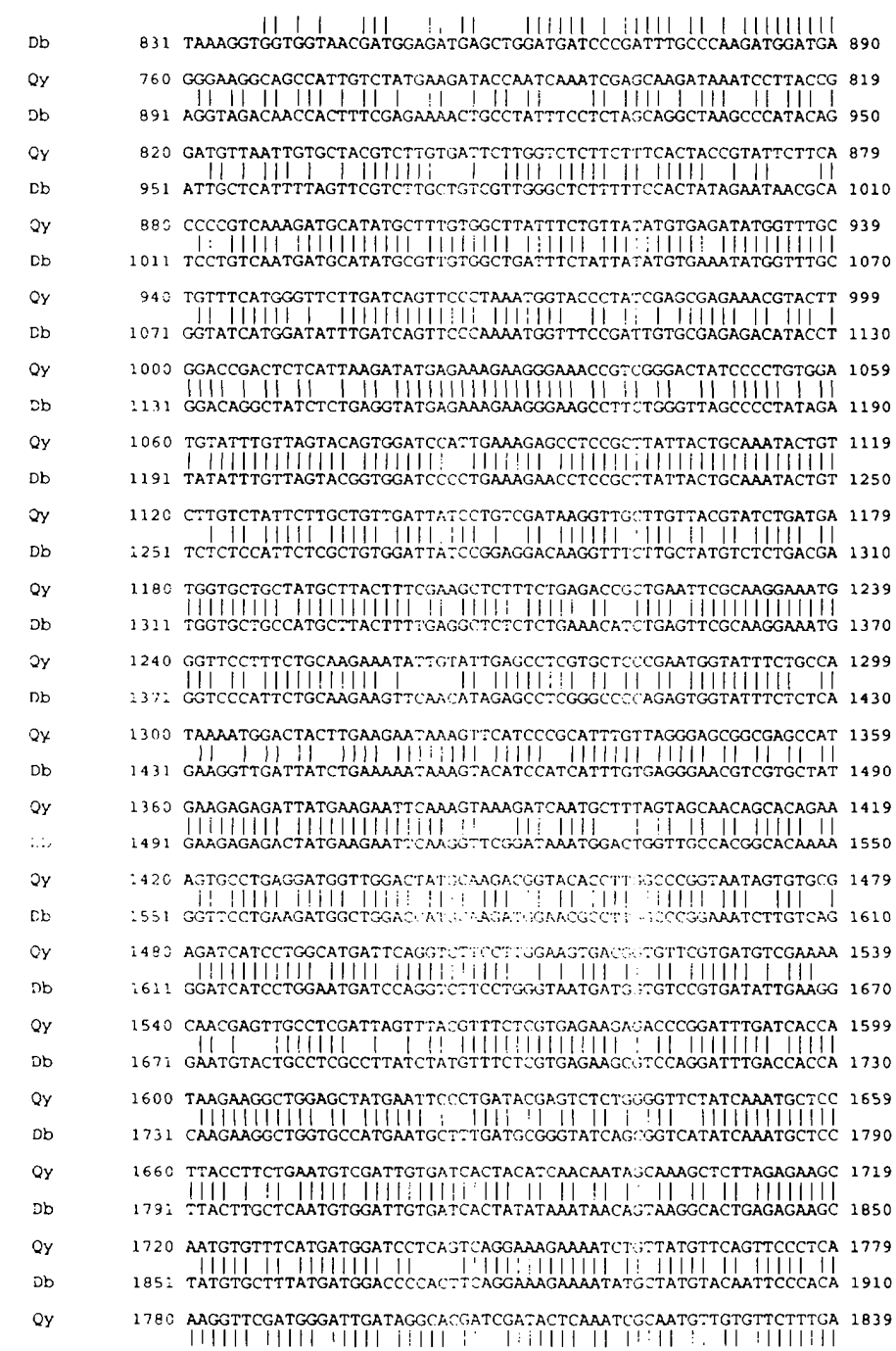
Qy        472 GGACGTTGAGATTTCTTCTGATAGACATGCTCTTATTGTTCCTCCTTCACTTGGTGGTCA 531
              ||| | |   |||||  |||| |  |||||||||  |||| | ||  |   ||||
Db        597 GGATGATACAATTTCCGCTGACAAGCATGCTCTTATCATTCCACCGT---TTATGGGTCG 653

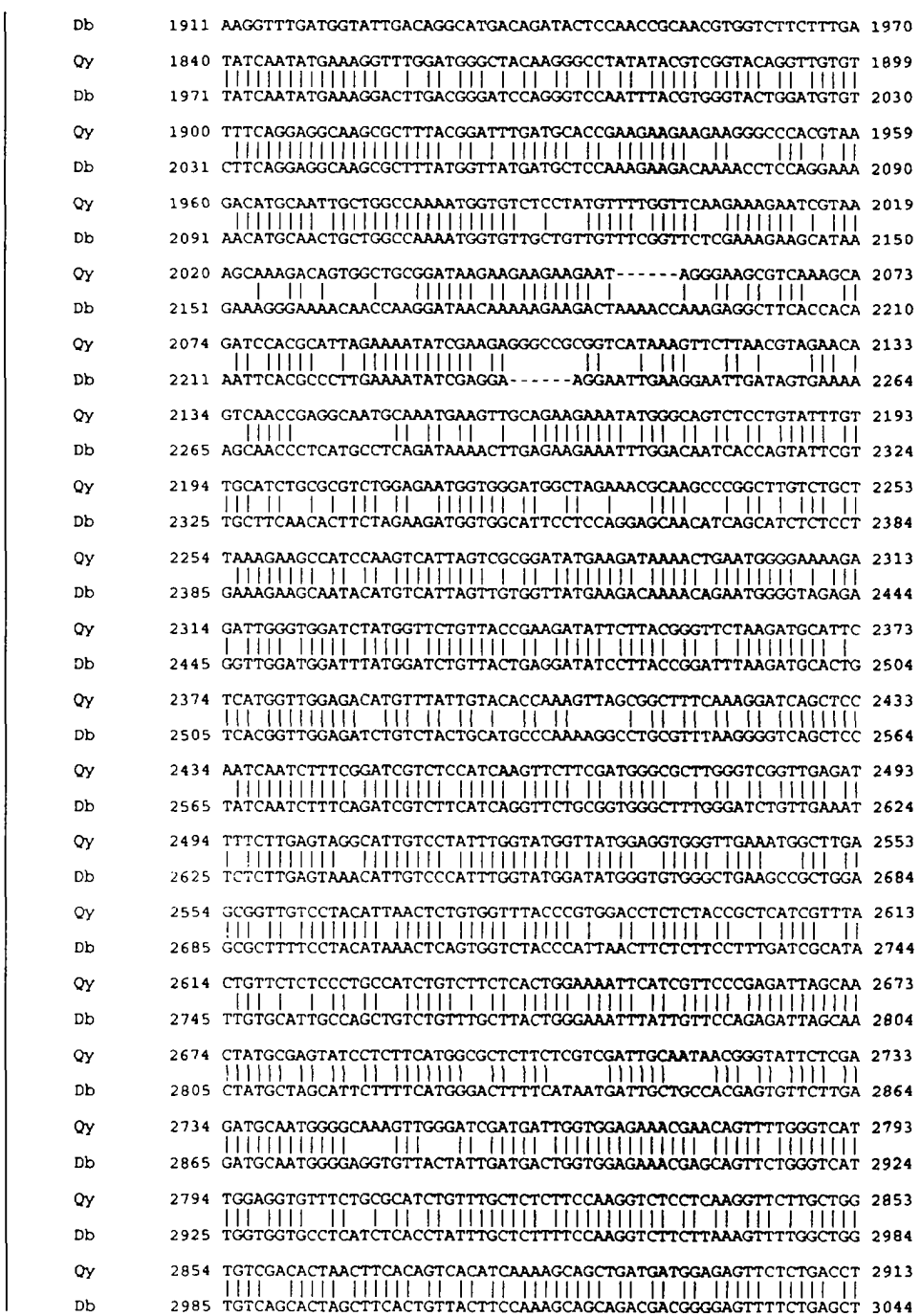
Qy        532 TGGCAATAGAGTTCATCCTGTTTCTCTTTCTGACCCGACCGTGGCTGCACATCGAAGGCT 591
              || || | ||||||||||||  || |||||  |  || || ||  | |  | |
Db        654 CGGAAAGAAAGTTCATCCTGTGCCTTATTCTGA---TTCCATGTCTTTGCCCCCTCGACC 710

Qy        592 GATGGTACCTCAGAAAGATCTTGCGGTTTATGGTTATGGAAGTGTCGCTTGGAAAGATCG 651
              |||| ||| |||||||| | || |||||||  ||||| | ||| || |||||  |  |
Db        711 CATGGATCCTAAGAAAGATTTGGCAGTTTATGGATATGGGACTGTTGCATGGAAGGAAAG 770

Qy        652 GATGGAGGAATGGAAGAGAAAGCAGAATGAGAAACTTCAGGTTGTTAGGCATGAA----- 706
              |||||||| |||||| ||| |||||||| || | |||| ||| ||| | |
Db        771 AATGGAGGACTGGAAGAAAAAACAGAATGATAAATTACAGGTGGTTAAGCACGGAGGCGG 830

Qy        707 -------GGAGATCCTGATTTTGAAGATGGTGATGATGCTGATTTTCCAATGATGGATGA 759
```





```
y       2914 TTACCTCTTCAAATGGACTTCACTTCTCATCCCTCCAATGACTCTACTCATCATAAACGT 2973
             ||| || ||||| ||||| ||  |  | || ||||| |||||| |  | || || ||| |
b       3045 TTATCTGTTCAAGTGGACATCTTTGTTGATTCCTCCTATGACTTTGTTAATTATTAACAT 3104

y       2974 CATTGGAGTCATAGTCGGAGTCTCTGATGCCATCAGCAATGGATACGACTCGTGGGGACC 3033
             |||||||||||||||| ||| | || ||||| |||| ||| || || || || ||||| ||
b       3105 CATTGGAGTCATAGTTGGAATTTCGGATGCAATCAACAACGGTTATGATTCTTGGGGTCC 3164

y       3034 GCTTTTCGGAAGACTGTTCTTTGCACTTTGGGTCATCATTCATCTTTACCCGTTCCTTAA 3093
             || ||||| || || ||||| ||  | |||||||| ||||||| || || ||||| ||
b       3165 TCTATTCGGTAGGCTTTTCTTCGCCTTGTGGGTCATTGTTCATCTATATCCCTTCCTCAA 3224

y       3094 AGGTTTGCTTGGGAAACAAGATAGAATGCCAACCATTATTGTCGTCTGGTCCATCCTCCT 3153
             |||| || | || | |||| ||| | | |||||||| ||||| |||||||| ||||| ||
b       3225 AGGTGTGATGGGCAGACAAAATAAAGTTCCAACCATCATTGTTGTCTGGTCTATCCTTCT 3284

y       3154 GGCCTCGATTCTTACACTTCTTTGGGTCCGGGTTAATCCGTTTGTGGCGAAAGGCGGTCC 3213
             ||| || ||  || |  |  | |||||||| || || || ||   ||| | ||| || |
b       3285 GGCATCTATCTTTTCCTTATTGTGGGTCCGAGTCAACCCATTCACGGCAAGAGGTGGACT 3344

y       3214 TATTCTCGAGATCTGTGGTTTAGACTGCTTGTGATT 3249
             || | ||  | || || || ||||||  || | |
b       3345 CGTTTTAGAAGTATGCGGGTTGGACTGCGAGTAAGT 3380


ESULT 10
F200531
OCUS       AF200531               3968 bp    mRNA    linear   PLN 31-AUG-2000
EFINITION  Zea mays cellulose synthase-7 (CesA-7) mRNA, complete cds.
CCESSION   AF200531
ERSION     AF200531.1  GI:9622885
EYWORDS    .
OURCE      Zea mays
 ORGANISM  Zea mays
           Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
           Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
           clade; Panicoideae; Andropogoneae; Zea.
EFERENCE   1  (bases 1 to 3968)
 AUTHORS   Holland,N., Holland,D., Helentjaris,T., Dhugga,K.S.,
           Xoconostle-Cazares,B. and Delmer,D.P.
 TITLE     A comparative analysis of the plant cellulose synthase (CesA) gene
           family
 JOURNAL   Plant Physiol. 123 (4), 1313-1324 (2000)
 MEDLINE   20398328
  PUBMED   10938350
EFERENCE   2  (bases 1 to 3968)
 AUTHORS   Dhugga,K.S. and Helentjaris,T.G.
 TITLE     Direct Submission
 JOURNAL   Submitted (01-NOV-1999) Agronomic Traits, Pioneer Hi-Bred
           International, Inc., 7250 NW 62nd Avenue, Johnston, IA 50131, USA
EATURES             Location/Qualifiers
     source         1. .3968
                    /organism="Zea mays"
                    /mol_type="mRNA"
                    /db_xref="taxon:4577"
                    /note="identified and isolated from the maize genome
                    database cDNA collection of Pioneer Hi-Bred International,
                    Inc"
     gene           1. .3968
                    /gene="CesA-7"
     CDS            143. .3403
                    /gene="CesA-7"
                    /note="CESA-7"
                    /codon_start=1
                    /product="cellulose synthase-7"
                    /protein_id="AAF89967.1"
                    /db_xref="GI:9622886"
                    /translation="MEASAGLVAGSHNRNELVVIRRDGDPGPKPPREQNGQVCQICGD
                    DVGLAPGGDPFVACNECAFPVCRDCYEYERREGTQNCPQCKTRYKRLKGCQRVTGDEE
                    EDGVDDLDNEFNWDGHDSQSVAESMLYGHMSYGRGGDPNGAPQAPQLNPNVPLLTNGQ
                    MVDDIPPEQHALVPSFMGGGGKRIHPLPYADPSLPVQPRSMDPSKDLAAYGYGSVAWK
                    ERMENWKQRQERMHQTGNDGGGDDGDDADLPLMDEARQQLSRKIPLPSSQINPYRMII
```

```
                    IIRLVVLGFFFHYRVMHPVNDAFALWLISVICEIWFAMSWILDQFPKWFPIERETYLD
                    RLSLRFDKEGQPSQLAPIDFFVSTVDPLKEPPLVTTNTVLSILSVDYPVDKVSCYVSD
                    DGAAMLTFEALSETSEFAKKWVPFCKRYNIEPRAPEWYFQQKIDYLKDKVAANFVRER
                    RAMKREYEEFKVRINALVAKAQKVPEEGWTMQDGTPWPGNNVRDHPGMIQVFLGQSGG
                    LDCEGNELPRLVYVSREKRPGYNHHKKAGAMNALVRVSAVLTNAPYLLNLDCDHYINN
                    SKAIKEAMCFMMDPLLGKKVCYVQFPQRFDGIDRHDRYANRNVVFFDINMKGLDGIQG
                    PIYVGTGCVFRRQALYGYDAPKTKKPPSRTCNCWPKWCFCCCCFGNRKQKKTTKPKTE
                    KKKLLFFKKEENQSPAYALGEIDEAAPGAENEKAGIVNQQKLEKKFGQSSVFVTSTLL
                    ENGGTLKSASPASLLKEAIHVISCGYEDKTDWGKEIGWIYGSVTEDILTGFKMHCHGW
                    RSIYCIPKRVAFKGSAPLNLSDRLHQVLRWALGSIEIFFSNHCPLWYGYGGGLKFLER
                    FSYINSIVYPWTSIPLLAYCTLPAICLLTGKFITPELNNVASLWFMSLFICIFATSIL
                    EMRWSGVGIDDWWRNEQFWVIGGVSSHLFAVFQGLLKVIAGVDTSFTVTSKGGDDEEF
                    SELYTFKWTTLLIPPTTLLLLNFIGVVAGVSNAINNGYESWGPLFGKLFFAFWVIVHL
                    YPFLKGLVGRQNRTPTIVIVWSILLASIFSLLWVRIDPFLAKDDGPLLEECGLDCN"
BASE COUNT     1022 a    837 c   1018 g   1091 t
ORIGIN

  Query Match          4: .%;  Score 1489.8;  DB 8;  Length 3968;
  Best Local Similarity   6: .%;  Pred. No. 0;
  Matches 2189;  Conservative    0;  Mismatches  1032;  Indels   36;  Gaps    5;

Qy          9 GGTTAATCGCCGGTTCTCACAACAGGAATGAGTTTGTCCTCATTAATGCCGATGAGAATG 68
              || |  | ||||: || |||||| | ||  || | ||| ||||       ||| :   ||
Db        159 GGCTGGTGGCCGGCTCCCACAACCGCAACGAGCTCGTCGTCATCCGCCGCGACGGCGATC 218

Qy         69 CCCGAATAAGATCAGTCCAAGAGCTGAGTGGACAGACATGTCAAATCTGCAGAGATGAGA 128
              ||  |    |   |  ! |||| ||  || |'|   || || || ||| | || ||
Db        219 CCGGGCCGAAGCCGCCGCGGGAGCAGAACGGGCAGGTGTGCCAGATTTGCGGCGACGACG 278

Qy        129 TCGAATTGACTGTTGATGGAGAACCGTTTGTGGCATGTAACGAATGTGCATTCCCTGTGT 188
              |||  |  |    |  || || || || ||||| || ||||| || || ||||| || |
Db        279 TCGGCCTTGCCCCCGGCGGGGACCCCTTCGTGGGGTGCAACGAGTGCGCCTTCCCCGTCT 338

Qy        189 GTAGACCTTGCTATGAGTACGAAAGACGAGAAGGCAATCAAGCTTGTCCACAGTGCAAAA 248
              |  |     ||||| |: ||||  | || || ' || ||    || || ||||||| |
Db        339 GCCGGGACTGCTACGAATACGAGCGCCGGGAGGGCACGCAGAACTGCCCCCAGTGCAAGA 398

Qy        249 CCCGTTTCAAACGTCTTAAAGGAAGTCCAAGAGTTGAAGGTGATGAAGAGGAAGATGACA 308
              | || | ||| || || || ||  | : | | ||    ||||| || ||||| || | |
Db        399 CTCGATACAAGCGCCTCAAGGGCTGCCAACGTGTGACCGGTGACGAGGAGGAGGACGGCG 458

Qy        309 TTGATGATTTAGACAATGAGTTTGA- -GTATGGAAATAATGGGATTGGATTTGATCAGG 365
              | |||||  | ||||| |||||  |  | | |:  || |   |     | |  ||
Db        459 TCGATGACCTGGACAACGAGTTCAACTGGGACGGCCATGACTCGCAGTCTGTGGCCGAGT 518

Qy        366 TTTCTGAAGGTATGTCAATCTCTCGTCGCAACTCCGGTTTCCCACAATCTGATTTGGATT 425
                                ||          ||| |||  |    |       |
Db        519 CCATGCTCTACGGCCACATGAGCTACGGCCGTGGAGGTGACCCTAATGGCGCGCCACAAG 578

Qy        426 CAGCTCCACCTGGCTCTCAGATTCCATTGCTGACTTACGGCGACGAGGACGTTGAGATTT 485
              |    |  |    | |  | ||||| | || ||  ||||  |   ||  | ||| ||
Db        579 CTTTCCAGCTCAACCCCAATGTTCCACTCCTCACCAACGGGCAAATGGTGGATGACATCC 638

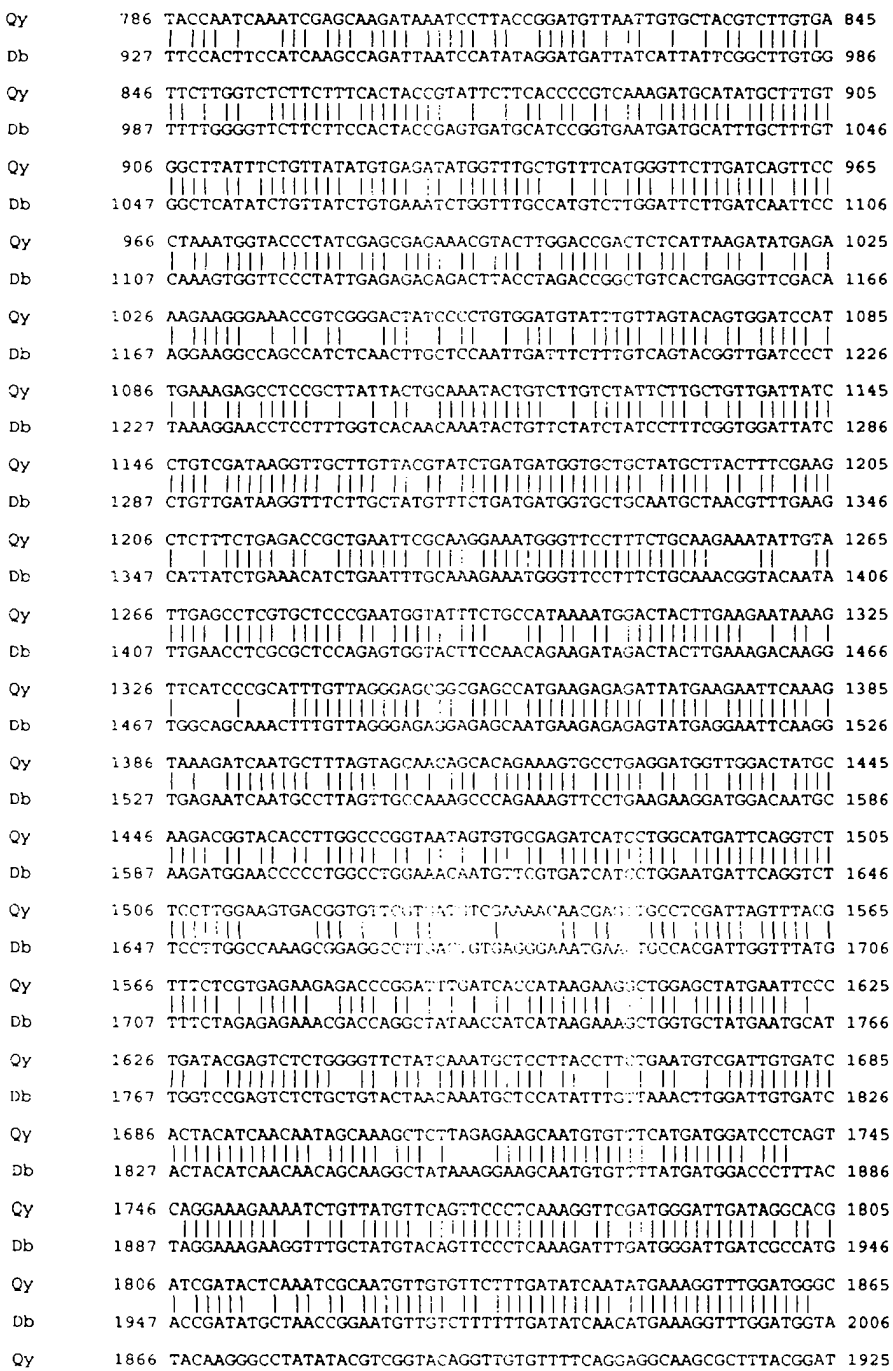
Qy        486 CTTCTGATAGACATGCTCTTATTGTTCCTCCTTCACTTGGTGGTCATGGCAATAGAGTTC 545
              | | ||    || || |  : || ||| ||:    | ||||||    || || ||  : |
Db        639 CACCGGAGCAGCACGCGC---TGGTGCCTTCTTTCATGGGTGGTGGGGGAAAGAGGATAC 695

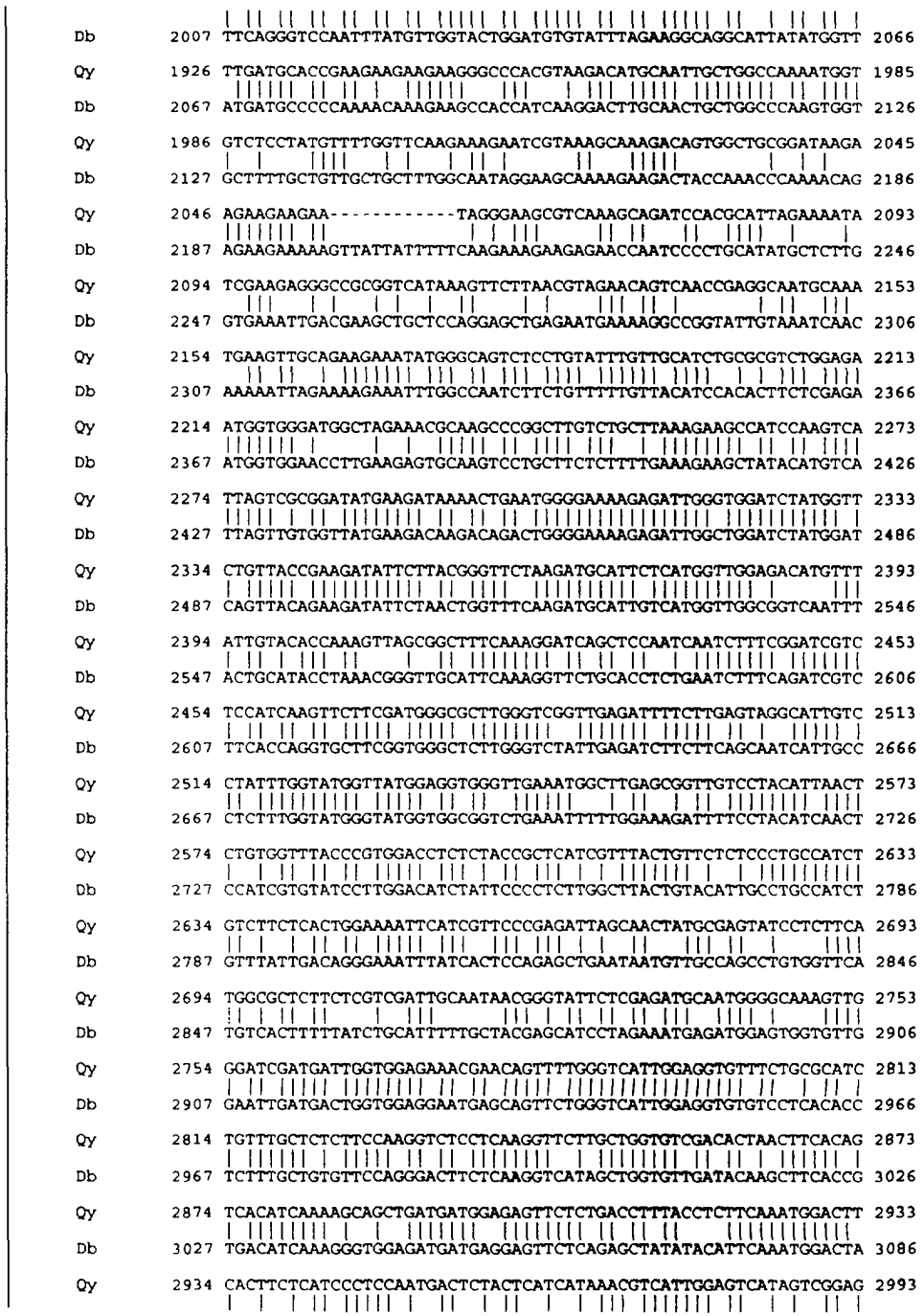
Qy        546 ATCCTGTTTCTCTTTCTGACCCGACCGTGGCTGCACATCGAAGGCTGATGGTACCTCAGA 605
              ||||  || ||  | ! || || | | |  !'|  || |  |||   |||| ||    |
Db        696 ATCCCCTTCCTTATGCGGATCCCAGCTTACCTGTGCAACCCAGGTCTATGGACCCATCCA 755

Qy        606 AAGATCTTGCGGTTTATGGTTATGGAAGTGTCGCTTGGAAAGATCGGATGGAGGAATGGA 665
              | |||||||| |  |||| |||| |||| |||||||| || |||||||||  | ||||
Db        756 AGGATCTTGCTGCATATGGGTATGGTAGTGTTGCTTGGAAGGAACGGATGGAGAATTGGA 815

Qy        666 AGAGAAAGCAGAATGAGAAACTTCAGGTTGTTAGGCATGAAGGAGATCCTGATTTTGAAG 725
              ||   |  ||  |   ||  |  |||   |  |          |||  || |  ||| |
Db        816 AGCAGAGACAAGAGAGGATGCACCAGACGGGGA---------ATGATGGTGGTGGTGATG 866

Qy        726 ATGGTGATGATGCTGATTTTCCAATGATGGATGAGGGAAGGCAGCCATTGTCTATGAAGA 785
              ||||||| |||||||| | ||| | |||||||| | ||| || | | |||| | ||| |
Db        867 ATGGTGACGATGCTGATCTACCACTAATGGATGAAGCAAGACAACAACTGTCCAGGAAAA 926
```





```
b       3087 CCTTATTGATACCTCCTACCACCTTGCTTCTATTGAACTTCATTGGTGTGGTCGCTGGCG 3146

y       2994 TCTCTGATGCCATCAGCAATGGATACGACTCGTGGGGACCGCTTTTCGGAAGACTGTTCT 3053
             | ||  |||| ||||  || |||||  || || |||||  || || || || |  || ||||
b       3147 TTTCAAATGCGATCAATAACGGATATGAGTCATGGGGCCCCCTCTTTGGGAAGCTATTCT 3206

y       3054 TTGCACTTTGGGTCATCATTCATCTTTACCCGTTCCTTAAAGGTTTGCTTGGGAAACAAG 3113
             ||||| |||||||  ||  | ||||||| || || || |||||||| |||| |  |||
b       3207 TTGCATTTTGGGTGATTGTCCATCTTTATCCCTTTCTCAAAGGTTTGGTTGGAAGGCAAA 3266

y       3114 ATAGAATGCCAACCATTATTGTCGTCTGGTCCATCCTCCTGGCCTCGATTCTTACACTTC 3173
             | || |  ||||| ||| |  |||||||||||| || ||||| || ||  |  | || |
b       3267 ACAGGACACCAACGATTGTCATCGTCTGGTCCATTCTGCTGGCTTCAATCTTCTCGCTCC 3326

y       3174 TTTGGGTCCGGGTTAATCCGTTTGTGGCGA---AAGGCGGTCCTATTCTCGAGATCTGTG 3230
             ||||||| ||| || |||| ||  | ||||   | |  |||||  |||| |||   ||||
b       3327 TTTGGGTTCGGATTGATCCTTTCCTTGCGAAGGATGATGGTCCGCTTCTTGAGGAGTGTG 3386

y       3231 GTTTAGACTGC 3241
             |||| || |||
b       3387 GTTTGGATTGC 3397
```

```
ESULT 11
F200530
OCUS       AF200530               3538 bp    mRNA    linear   PLN 31-AUG-2000
EFINITION  Zea mays cellulose synthase-6 (CesA-6) mRNA, complete cds.
CCESSION   AF200530
ERSION     AF200530.1  GI:9622883
EYWORDS    .
OURCE      Zea mays
 ORGANISM  Zea mays
           Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
           Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
           clade; Panicoideae; Andropogoneae; Zea.
EFERENCE   1  (bases 1 to 3538)
 AUTHORS   Holland,N., Holland,D., Helentjaris,T., Dhugga,K.S.,
           Xoconostle-Cazares,B. and Delmer,D.P.
 TITLE     A comparative analysis of the plant cellulose synthase (CesA) gene
           family
 JOURNAL   Plant Physiol. 123 (4), 1313-1324 (2000)
 MEDLINE   20398328
  PUBMED   10938350
EFERENCE   2  (bases 1 to 3538)
 AUTHORS   Dhugga,K.S. and Helentjaris,T.G.
 TITLE     Direct Submission
 JOURNAL   Submitted (01-NOV-1999) Agronomic Traits, Pioneer Hi-Bred
           International, Inc., 7250 NW 62nd Avenue, Johnston, IA 50131, USA
EATURES             Location/Qualifiers
     source         1. .3538
                    /organism="Zea mays"
                    /mol_type="mRNA"
                    /db_xref="taxon:4577"
                    /note="identified and isolated from the maize genome
                    database cDNA collection of Pioneer Hi-Bred International,
                    Inc"
     gene           1. .3538
                    /gene="CesA-6"
     CDS            43. .3222
                    /gene="CesA-6"
                    /note="CESA-6"
                    /codon_start=1
                    /product="cellulose synthase-6"
                    /protein_id="AAF89966.1"
                    /db_xref="GI:9622884"
                    /translation="MDQRNGQVCQICGDDVGRNPDGEPFVACNECAFPICRDCYEYER
                    REGTQNCPQCKTRFKRFKGCARVPGDEEEDGVDDLENEFNWSDKHDSQYLAESMLHAH
                    MSYGRGADLDGVPQPFHPIPNVPLLTNGQMVDDIPPDQHALVPSFVGGGGKRIHPLPY
                    ADPNLPVQPRSMDPSKDLAAYGYGSVAWKERMESWKQKQERMHQTRNDGGGDDGDDAD
                    LPLMDEARQPLSRKIPLPSSQINPYRMIIIIRLVVLCFFFHYRVMHPVPDAFALWLIS
                    VICEIWFAMSWILDQFPKWPPIERETYLDRLSLRFDKEGHPSQLAPVDFFVSTVDPLK
                    EPPLVTANTVLSILSVDYPVDKVSCYVSDDGAAMLTFEALSETSEFAKKWVPFCKRYS
                    LEPRAPEWYFQQKIDYLKDKVAPNFVRERRAMKREYEEFKVRINALVAKAQKVPEEGW
                    TMQDGTPWPGNNVRDHPGMIQVFLGQSGGHDVEGNELPRLVYVSREKRPGYNHHKKAG
                    AMNALVRVSAVLTNAPYLLNLDCDHYINNSKAIKEAMCFMMDPLLGKKVCYVQFPQRF
                    DGIDRHDRYANRNVVFFDINMKGLDGIQGPIYVGTGCVFRRQALYGYDAPKTKKPPSR
                    TCNCWPKWCICCCCFGNRKTKKKTKTSKPKFEKIKKLFKKKENQAPAYALGEIDEAAP
                    GAENEKASIVNQQKLEKKFGQSSVFVASTLLENGGTLKSASPASLLKEAIHVISCGYE
                    DKTGWGKDIGWIYGSVTEDILTGFKMHCHGWRSIYCIPKRAAFKGSAPLNLSDRFHQV
                    LRWALGSIEILFSNHCPLWYGYGGGLKFLERFSYINSIVYPWTSIPLLAYCTLPAICL
                    LTGKFITPELNNVASLWFMSLFICIFATSILEMRWSGVGIDDWWRNEQFWVIGGVSSH
                    LFAVFQGLLKVIAGVDTSFTVTSKGSDDEEFSELYTFKWTTLLIPPTTLLLLNFIGVV
                    AGISNAINNGYESWGPLFGKLFFAFWVIVHLYPFLKGLVGRQNRTPTIVIVWSILLAS
                    IFSLLWVRIDPFLAKDDGPLLEECGLDCN"
BASE COUNT      876 a    813 c    928 g    920 t      1 others
ORIGIN
```

```
  Query Match           42.5%;  Score 1462.4;  DB 8;  Length 3538;
  Best Local Similarity 67.3%;  Pred. No. 0;
  Matches 2174;  Conservative    0;  Mismatches 1011;  Indels   45;  Gaps    6;

Qy        42 TTGTCCTCATTAATGCCGATGAGAATGCCCGAATAAGATCAGTCCAAGAGCTGAGTGGAC 101
             | ||| ||||      ||||  |  |  |    |  |  |  | ||| ||  || |
Db         2 TCGTCGTCATCCGCTGCGATGGCGAGCCAGGTCGGAAGCCCATGGACCAGCGGAACGGCC 61

Qy       102 AGACATGTCAAATC CAGAGATGAATCGAATCGAACTGTTGATGGAGAACCGTTTGTGG 161
             ||  ||  |  ||  ||  |  |  |  |  |      ||  ||  ||  ||  ||||
Db        62 AGGTGTGCCAGATTTGCGGCGACGACGTGGGCTGCAACCCCGACGGGGAGCCTTTCGTGG 121

Qy       162 CATGTAACGAATGTGCATTCCCTGTGTGTAGACCTTGCTATGAGTACGAAAGACGAGAAG 221
             | || ||||| || ||  |||||  |  ||  |  |||||  ||||||||  |  ||  || |
Db       122 CCTGCAACGAGTGCGCCTTCCCCATCTGCCGGGACTGCTACGAGTACGAGCGCCGCGAGG 181

Qy       222 GCAATCAAGCTTGTCCACAGTGCAAAACCCGTTTCAAACGTCTTAAAGGAAGTCCAAGAG 281
             |||  ||    ||  |  ||||||||  |||||  ||||  ||  |  ||  ||  |  |
Db       182 GCACGCAGAACTGCCCCCAGTGCAAGACCCGCTTCAAGCGCTTCAAGGGGTGCGCGCGCG 241

Qy       282 TTGAAGGTGATGAAGAGGAAGATGACATTGATGATTTAGACAATGAGTTTGAGTATGGAA 341
             |    ||  ||  ||  |||||  ||  |  |  |  ||  |  ||  ||  ||||  |  |    |
Db       242 TGCCCGGGGACGAGGAGGAGGACGGCGTCGACGACCTGGAGAACGAGTTCAACTGGAGCG 301

Qy       342 ATAATGGGATTGGATTTGATCAGGTTTCTGAAGGTATGTCAATCTCTCGTCGCAACTCCG 401
             | ||   |  |  |  |  |  |  ||   |||    |  |  |    |  ||  ||
Db       302 ACAA---GCACGACTCCCAGTACCTCGCCGAGTCCATGCTCCACGCCCACATGAGCTACG 358

Qy       402 G---------TTTCCCACAATCTGATTTGGATTCAGCTCCACCTGGCTCTCAGATTCCAT 452
             |         ||  |  |  |  |  |  ||  |  |  |  |  ||||
Db       359 GCCGCGGCGCCGACCTCGACGGCGTGCCGCAGCCATTCCACCCCATCCCCAATGTTCCCC 418

Qy       453 TGCTGACTTACGGCGACGAGGACGTTGAGATTTCTTCTGATAGACATGCTCTTATTGTTC 512
             | ||  ||  ||||  |  ||  ||  |||  ||  |  |  ||    ||  ||  |  ||||  |
Db       419 TCCTCACCAACGGACAGATGGTCGATGACATCCCGCCGGACCAGCACGCCC---TTGTGC 475

Qy       513 CTCCTTCACTTGGTGGTCATGGCAATAGAGTTCATCCTGTTTCTCTTTCTGACCCGACCG 572
             |  |  |  |  ||||  |  ||  ||  |||  |  |  |  ||  ||  |  |
Db       476 CCTCGTTCGTGGGTGGCGGGGGGAAGAGGATTCACCCTCTCCCGTACGCGGATCCCAACC 535

Qy       573 TGGCTGCACATCGAAGGCTGATGGTACCTCAGAAAGATCTTGCGGTTTATGGTTATGGAA 632
             |  |||  ||  |  |||  ||||  |||  ||  |||||  ||  |  |||||  ||  ||  |
Db       536 TTCCTGTGCAACCGAGGTCTATGGACCCTTCCAAGGATCTCGCCGCATATGGCTACGGGA 595

Qy       633 GTGTCGCTTGGAAAGATCGGATGGAGGAATGGAAGAGAAAGCAGAATGAGAAACTTCAGG 692
             | ||  ||  |||||  |  ||||||||  ||||||  ||||||  |  ||  |  |||
Db       596 GCGTAGCATGGAAGGAGAGGATGGAGAGCTGGAAGCAGAAGCAGGAGAGGATGCACCAGA 655

Qy       693 TTGTTAGGCATGAAGGAGATCCTGATTTTGAAGATGGTGATGATGCTGATTTTCCAATGA 752
             |   |||  ||  |           ||  ||||||||||||||  |||  |  |||  |  |
Db       656 CGAGGAACGATGGCGGCG---  -----GCGATGATGGTGATGATGCAGATCTACCACTAA 706

Qy       753 TGGATGAGGGAAGGCAGCCATTGTCTATGAAGATACCAATCAAATCGAGCAAGATAAATC 812
             |||||||  |  ||  ||||||||||||  |  |||||  ||  |  ||  |||  |  ||  ||  |
Db       707 TGGATGAAGCTAGACAGCCATTGTCCAGAAAGATCCCGCTTCCTTCAAGCCAAATCAACC 766

Qy       813 CTTACCGGATGTTAATTGTGCTACGTCTTGTGATTCTTGGTCTCTTCTTTCACTACCGTA 872
             | ||  |||||  |  ||  |  |  ||  ||  |||  ||  |  ||  |||||||  ||||||||
Db       767 CCTATAGGATGATTATAATAATTCGGCTAGTGGTTTTGTGTTTCTTCTTCCACTACCGAG 826
```



```
                 |||| || || ||||| || || ||||| |  || |||||||||| ||||||| ||| | |
b          2987 GATATGAATCATGGGGCCCCCTGTTCGGGAAGCTCTTCTTTGCATTTTGGGTGATCGTCC 3046

y          3075 ATCTTTACCCGTTCCTTAAAGGTTTGCTTGGGAAACAAGATAGAATGCCAACCATTATTG 3134
                |||||||||||||||| || ||| || |||||| ||  | || | |||||| ||| |
b          3047 ATCTTTACCCGTTCCTCAAGGGTCTGGTTGGGAGGCAGAACAGGACGCCAACGATTGTCA 3106

y          3135 TCGTCTGGTCCATCCTCCTGGCCTCGATTCTTACACTTCTTTGGGTCCGGGTTAATCCGT 3194
                | |||||||||||||||||||| ||||| |  | || |||||||||||| |  | ||||
b          3107 TTGTCTGGTCCATCCTCCTGGCTTCGATCTTCTCGCTGCTTTGGGTCCGGATCGACCCGT 3166

y          3195 TTGTGGCGA---AAGGCGGTCCTATTCTCGAGATCTGTGGTTTAGACTGC 3241
                |  | ||||   | |  |||||  |  | |||   |||||| | || |||
b          3167 TCCTTGCGAAGGATGATGGTCCCCTGTTGGAGGAGTGTGGTCTGGATTGC 3216


ESULT 12
F375439
OCUS       AF375439               1492 bp    mRNA    linear   PLN 22-FEB 2002
EFINITION  Arabidopsis thaliana AT5g64740/MVP7_7 mRNA, complete cds
CCESSION   AF375439
ERSION     AF375439.1  GI:14030696
EYWORDS    FLI_CDNA.
OURCE      Arabidopsis thaliana (thale cress)
 ORGANISM  Arabidopsis thaliana
           Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
           Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
           rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
EFERENCE   1  (bases 1 to 1492)
 AUTHORS   Cheuk,R., Chen,H., Kim,C.J., Koesema,E., Meyers,M.C., Shinn,P.,
           Banh,J., Bowser,L., Carninci,P., Chung,M.K., Goldsmith,A.D.,
           Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G.,
           Kawai,J., Lam,B., Lee,J.M., Lin,J., Liu,S.X., Miranda,M.,
           Narusaka,M., Nguyen,M., Palm,C.J., Pham,P.K., Quach,H.L.,
           Sakano,H., Sakurai,T., Satou,M., Seki,M., Southwick,A., Tang,C.C.,
           Toriumi,M., Yamada,K., Yu,G., Shinozaki,K., Davis,R.W.,
           Theologis,A. and Ecker,J.R.
 TITLE     Arabidopsis cDNA clones
 JOURNAL   Unpublished
EFERENCE   2  (bases 1 to 1492)
 AUTHORS   Cheuk,R., Chen,H., Kim,C.J., Koesema,E., Meyers,M.C., Shinn,P.,
           Banh,J., Bowser,L., Carninci,P., Chung,M.K., Goldsmith,A.D.,
           Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G.,
           Kawai,J., Lam,B., Lee,J.M., Lin,J., Liu,S.X., Miranda,M.,
           Narusaka,M., Nguyen,M., Palm,C.J., Pham,P.K., Quach,H.L.,
           Sakano,H., Sakurai,T., Satou,M., Seki,M., Southwick,A., Tang,C.C.,
           Toriumi,M., Yamada,K., Yu,G., Shinozaki,K., Davis,R.W.,
           Theologis,A. and Ecker,J.R.
 TITLE     Direct Submission
 JOURNAL   Submitted (01-MAY-2001) Salk Institute Genomic Analysis Laboratory
           (SIGnAL), Plant Biology Laboratory, The Salk Institute for
           Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037,
           USA
OMMENT     RIKEN Genomic Sciences Center (GSC) members carried out the
           collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN
           Arabidopsis Full-Length cDNA') : Seki,M., Narusaka,M., Ishida,J.,
           Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J.,
           Hayashizaki,Y. and Shinozaki,K.

           The Salk, Stanford, PGEC (SSP) Consortium members carried out the
           sequencing and annotation of the RAFL cDNAs: Shinn,P., Chen,H.,
           Cheuk,R., Kim,C.J., Koesema,E., Meyers,M.C., Tracy,S.E., Banh,J.
           Bowser,L., Chung,M.K., Goldsmith,A.D., Jones,T., Karlin-Neumann,G.,
           Lam,B., Lee,J.M., Lin,J., Liu,S.X., Miranda,M., Nguyen,M.,
           Palm,C.J., Pham,P.K., Quach,H.L., Sakano,H., Southwick,A.,
           Tang,C.C., Toriumi,M., Yamada,K., Yu,G., Davis,R.W., Theologis,A.,
           and Ecker,J.R.

           Shinn,P. (SSP/Salk) and Seki,M. (RIKEN GSC) contributed equally to
           this work.  Shinozaki,K. (RIKEN GSC) and Ecker,J.R. (SSP/Salk)
           contributed equally to this work as PIs.
EATURES              Location/Qualifiers
     source          1. .1492
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
                     /chromosome="5"
                     /clone="RAFL05-02-P19(R12458)"
                     /note="ecotype: Columbia"
     CDS             247. .1347
                     /note="cellulose synthase catalytic subunit"
                     /codon_start=1
                     /product="AT5g64740/MVP7_7"
                     /protein_id="AAK58023.1"
                     /db_xref="GI:14030697"
                     /translation="MQMKLEKKFGQSPVFVASARMENGGMARNASPACLLKEAIQVIS
                     CGYEDKTEWGKEIGWIYGSVTEDILTGFKMHSHGWRSVYCTPKLAAFEGSAPINLSDR
                     LHQVLRWALGSVEIFLSRHCPIWYGYGGGLKWLERLSYINSVVYPWTSLPLIVYCSLP
                     AICLLTGKFIVPEISNYASILFMALFSSIAITGILEMQWGKVGIDDWWRNEQFWVIGG
                     VSAHLFALFQGLLKVLAGVETNFTVTSKAADDGEFSDLYLFKWTSLLIPPMTLLIINV
                     IGVIVGVSNAISKGYDSWGPLFGRLFFALWVIIHLYPFLKGLLGKQDRMPTIIVVWSI
                     LLASIFSLWVAVNPFVAKGGPVLEVCGLGCD"
BASE COUNT      383 a    253 c    373 g    83 t
ORIGIN

  Query Match           42.8%;  Score 1454;  DB 8;  Length 1492;
  Best Local Similarity   98.9%;  Pred. No. 0;
  Matches 1476;  Conservative    0;  Mismatches   15;  Indels    1;  Gaps    1;

Qy       1900 TTTCAGGAGGCAAGCGCTTTACGGATTTGATGCACCGAAGAAGAAGAAGGGCCCACGTAA 1959
                |||||||||| ||| |||||||| | |||||  |||| |||||||||||||||| |||||  ||
Db          1 TTTCAGGAGGCAAGCGCTTTAC-GATTTGATGCACCGAAGAAGAAGAAGGGCCCACGTAA 59

Qy       1960 GACATGCAATTGCTGGCCAAAATGGTGTCTCGTATGTTTTGGTTCAAGAAAGAATCGTAA 2019
                |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db         60 GACATGCAATTGCTGGCCAAAATGGTGTCTCGTATGTTTTGGTTCAAGAAAGAATCGTAA 119

Qy       2020 AGCAAAGACAGTGGCTGCGGATAAGAAGAAGAAGAATAGGGAAGCGTCAAAGCAGATCCA 2079
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        120 AGCAAAGACAGTGGCTGCGGATAAGAAGAAGAAGAATAGGGAAGCGTCAAAGCAGATCCA 179

Qy       2080 CGCATTAGAAAAATATCGAAGAGGGCCGCGGTCATAAAGTTCTTAACGTAGAACAGTCAAC 2139
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        180 CGCATTAGAAAAATATCGAAGAGGGCCGCGTCACTAAAGGTTCTAACGTAGAACAGTCAAC 239

Qy       2140 CGAGGCAATGCAAATGAAGTTGCAGAAGAAATATGGGCAGTCTCCTGTATTTGTTGCATC 2199
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        240 CGAGGCAATGCAAATGAAGTTGGAGAAGAAATTTGGGCAGTCTCCTGTATTTGTTGCATC 299

Qy       2200 TGCGCGTCTGGAGAATGGTGGGATGGCTAGAAACGCAAGCCCGGCTTGTCTGCTTAAAGA 2259
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        300 TGCGCGTATGGAGAATGGTGGGATGGCTAGAAACGCAAGCCCGGCTTGTCTGCTTAAAGA 359

Qy       2260 AGCCATCCAAGTCATTAGTCGCGGATATGAAGATAAAACTGAATGGGGAAAAGAGATTGG 2319
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        360 AGCCATCCAAGTCATTAGTTGCGGATATGAAGATAAAACTGAATGGGGAAAAGAGATTGG 419

Qy       2320 GTGGATCTATGGTTCTGTTACCGAAGATATTCTTACGGGTTCTAAGATGCATTCTCATGG 2379
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        420 GTGGATCTATGGTTCTGTTACCGAAGATATTCTTACGGGTTTTAAGATGCATTCTCATGG 479

Qy       2380 TTGGAGACATGTTTATTGTACACCAAAGTTAGCGGCTTTCAAAGGATCAGCTCCAATCAA 2439
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        480 TTGGAGATCTGTTTATTGTACACCAAAGTTAGCGGCTTTCGAAGGATCAGCTCCAATCAA 539

Qy       2440 TCTTTCGGATCGTCTCCATCAAGTTCTTCGATGGGCGCTTGGGTCGGTTGAGATTTTCTT 2499
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        540 TCTTTCGGATCGTCTCCATCAAGTTCTTCGATGGGCGCTTGGGTCGGTTGAGATTTTCTT 599

Qy       2500 GAGTAGGCATTGTCCTATTTGGTATGGTTATGGAGGTGGGTTGAAATGGCTTGAGCGGTT 2559
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        600 GAGTAGGCATTGTCCTATTTGGTATGGTTATGGAGGTGGGTTGAAATGGCTTGAGCGGTT 659

Qy       2560 GTCCTACATTAACTCTGTGGTTTACCCGTGGACCTCTCTACCGCTCATCGTTTACTGTTC 2619
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        660 GTCCTACATTAACTCTGTGGTTTACCCGTGGACCTCTCTACCGCTCATCGTTTACTGTTC 719
```

```
Qy       2620 TCTCCCTGCCATCTGTCTTCTCACTGGAAAATTCATCGTTCCCGAGATTAGCAACTATGC 2679
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        720 TCTCCCTGCCATCTGTCTTCTCACTGGAAAATTCATCGTTCCCGAGATTAGCAACTATGC 779

Qy       2680 GAGTATCCTCTTCATGGCGCTCTTCTCGTCGATTGCAATAACGGGTATTCTCGAGATGCA 2739
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        780 GAGTATCCTCTTCATGGCGCTCTTCTCGTCGATTGCAATAACGGGTATTCTCGAGATGCA 839

Qy       2740 ATGGGGCAAAGTTGGGATCGATGATTGGTGGAGAAACGAACAGTTTTGGGTCATTGGAGG 2799
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        840 ATGGGGCAAAGTTGGGATCGATGATTGGTGGAGAAACGAACAGTTTTGGGTCATTGGAGG 899

Qy       2800 TGTTTCTGCGCATCTGTTTGCTCTCTTCCAAGGTCTCCTCAAGGTTCTTGCTGGTGTCGA 2859
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        900 TGTTTCTGCGCATCTGTTTGCTCTCTTCCAAGGTCTCCTCAAGGTTCTTGCTGGTGTCGA 959

Qy       2860 CACTAACTTCACAGTCACATCAAAAGCAGCTGATGATGGAGAGTTCTCTGACCTTTACCT 2919
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        960 CACTAACTTCACAGTCACATCAAAAGCAGCTGATGATGGAGAGTTCTCTGACCTTTACCT 1019

Qy       2920 CTTCAAATGGACTTCACTTCTCATCCCTCCAATGACTCTACTCATCATAAACGTCATTGG 2979
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1020 CTTCAAATGGACTTCACTTCTCATCCCTCCAATGACTCTACTCATCATAAACGTCATTGG 1079

Qy       2980 AGTCATAGTCGGAGTCTCTGATGCCATCAGCAATGGATACGACTCGTGGGGACCGCTTTT 3039
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1080 AGTCATAGTCGGAGTCTCTGATGCCATCAGCAATGGATACGACTCGTGGGGACCGCTTTT 1139

Qy       3040 CGGAAGACTGTTCTTTGCACTTTGGGTCATCATTCATCTTTACCCGTTCCTTAAAGGTTT 3099
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1140 CGGAAGACTGTTCTTTGCACTTTGGGTCATCATTCATCTTTACCCGTTCCTTAAAGGTTT 1199

Qy       3100 GCTTGGGAAACAAGATAGAATGCCAACCATTATTGTCGTCTGGTCCATCCTCCTGGCCTC 3159
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1200 GCTTGGGAAACAAGATAGAATGCCAACCATTATTGTCGTCTGGTCCATCCTCCTGGCCTC 1259

Qy       3160 GATTCTTACACTTCTTTGGGTCCGGGTTAATCCGTTTGTGGCGAAAGGCGGTCCTATTCT 3219
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1260 GATTCTTACACTTCTTTGGGTCCGGGTTAATCCGTTTGTGGCGAAAGGCGGTCCTATTCT 1319

Qy       3220 CGAGATCTGTGGTTTAGACTGCTTGTGATTCGATTGACCGGTGGATGGGTTGGTGAAAAA 3279
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1320 CGAGATCTGTGGTTTAGACTGCTTGTGATTCGATTGACCGGTGGATGGGTTGGTGAAAAA 1379

Qy       3280 GGTTTAATTCCCACGGATCAAAGAGAGGTAAGAGAGATATTGTTTTACCTCTAAAAGACT 3339
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1380 GGTTTAATTCCCACGGATCAAAGAGAGGTAAGAGAGATATTGTTTTACCTCTAAAAGACT 1439

Qy       3340 CCTTCATTGTGTTCATTAGATGATGAAAAATGAAAAGAAAAAGAAGATTTAA 3391
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1440 CCTTCATTGTGTTCATTAGATGATGAAAAATGAAAAGAAAAAGAAGATTTAA 1491
```

RESULT 13
AF200532

```
LOCUS       AF200532               3812 bp    mRNA    linear   PLN 31-AUG-2000
DEFINITION  Zea mays cellulose synthase-8 (CesA 8) mRNA, complete cds.
ACCESSION   AF200532
VERSION     AF200532.1  GI:9622887
KEYWORDS    .
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 3812)
  AUTHORS   Holland,N., Holland,D., Helentjaris,T., Dhugga,K.S.,
            Xoconostle-Cazares,B. and Delmer,D.P.
  TITLE     A comparative analysis of the plant cellulose synthase (CesA) gene
            family
  JOURNAL   Plant Physiol. 123 (4), 1313 1324 (2000)
  MEDLINE   20398328
  PUBMED    10938350
REFERENCE   2  (bases 1 to 3812)
  AUTHORS   Dhugga,K.S. and Helentjaris,T.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-NOV-1999) Agronomic Traits, Pioneer Hi-Bred
            International, Inc., 7250 NW 62nd Avenue, Johnston, IA 50131, USA
FEATURES             Location/Qualifiers
     source          1. .3812
                     /organism="Zea mays"
                     /mol_type="mRNA"
                     /db_xref="taxon:4577"
                     /note="identified and isolated from the maize genome
                     database cDNA collection of Pioneer Hi-Bred International,
                     Inc"
     gene            1. .3812
                     /gene="CesA-8"
     CDS             214. .3498
                     /gene="CesA-8"
                     /note="CESA-8"
                     /codon_start=1
                     /product="cellulose synthase-8"
                     /protein_id="AAF89968.1"
                     /db_xref="GI:9622888"
                     /translation="MEASAGLVAGSHNRNELVVIRRDRESGAAGGGAARRAEAPCQIC
                     GDEVGVGFDGEPFVACNECAFPVCRACYEYERREGSQACPQCRTRYKRLKGCPRVAGD
                     EEEDGVDDLEGEFGLQDGAAHEDDPQYVAESMLRAQMSYGRGGDAHPGFSPVPNVPLL
                     TNGQMVDDIPPEQHALVPSYMSGGGGGGKRIHPLPFADPNLPVQPRSMDPSKDLAAYG
                     YGSVAWKERMEGWKQKQERLQHVRSEGGGDWDGDDADLPLMDEARQPLSRKVPISSSR
                     INPYRMIIVIRLVVLGFFFHYRVMHPAKDAFALWLISVICEIWFAMSWILDQFPKWLP
                     IERETYLDRLSLRFDKEGQPSQLAPIDFFVSTVDPTKEPPLVTANTVLSILSVDYPVE
                     KVSCYVSDDGAAMLTFEALSETSEFAKKWVPFSKKFNIEPRAPEWYFQQKIDYLKDKV
                     AASFVRERRAMKREYEEFKVRINALVAKAQKVPEEGWTMQDGSPWPGNNVRDHPGMIQ
                     VFLGQSGGRDVEGNELPRLVYVSREKRPGYNHHKKAGAMNALVRVSAVLSNAAYLLNL
                     DCDHYINNSKAIKEAMCFMMDPLVGKKVCYVQFPQRFDGIDKNDRYANRNVVFFDINM
                     KGLDGIQGPIYVGTGCVFRRQALYGYDAPKTKKPPSRTCNCWPKWCLSCCCSRNKKK
                     KTTKPKTEKKKRLFFKKAENPSPAYALGEIDEGAPGADIEKAGIVNQQKLEKKFGQSS
                     VFVASTLLENGGTLKSASPASLLKEAIHVISCGYEDKTDWGKEIGWIYGSITEDILTG
                     FKMHCHGWRSIYCIPKRPAFKGSAPLNLSDRLHQVLRWALGSVEIFFSKHCPLWYGYG
                     GGLKFLERFSYINSIVYPWTSIPLLAYCTLPAICLLTGKFITPELTNVASIWFMALFI
                     CISVTGILEMRWSGVAIDDWWRNEQFWVIGGVSAHLFAVFQGLLKVFAGIDTSFTVTS
                     KAGDDEEFSELYTFKWTTLLIPPTTLLLLNFIGVVAGISNAINNGYESWGPLFGKLFF
                     AFWVIVHLYPFLKGLVGRQNRTPTIVIVWSILLASIFSLLWVRVDPFLAKSNGPLLEE
                     CGLDCN"
BASE COUNT      873 a    966 c   1111 g    862 t
ORIGIN
```

```
  Query Match             41.1%;  Score 1416.4;  DB 8;  Length 3812;
  Best Local Similarity   66.7%;  Pred. No. 0;
  Matches 2128;  Conservative    0;  Mismatches 1026;  Indels   36;  Gaps    6;

Qy        103 GACATGTCAAATCTGCAGAGATGAGATCGAATTGACTGTTGATGGAGAACCGTTTGTGGC 162
              | | || || || ||| | || ||| |||   ||    | || || || || || |||||
Db        330 GCCGTGCCAGATATGCGGCGACGAGGTCGGGGTGGGCTTCGACGGGGAGCCCTTCGTGGC 389

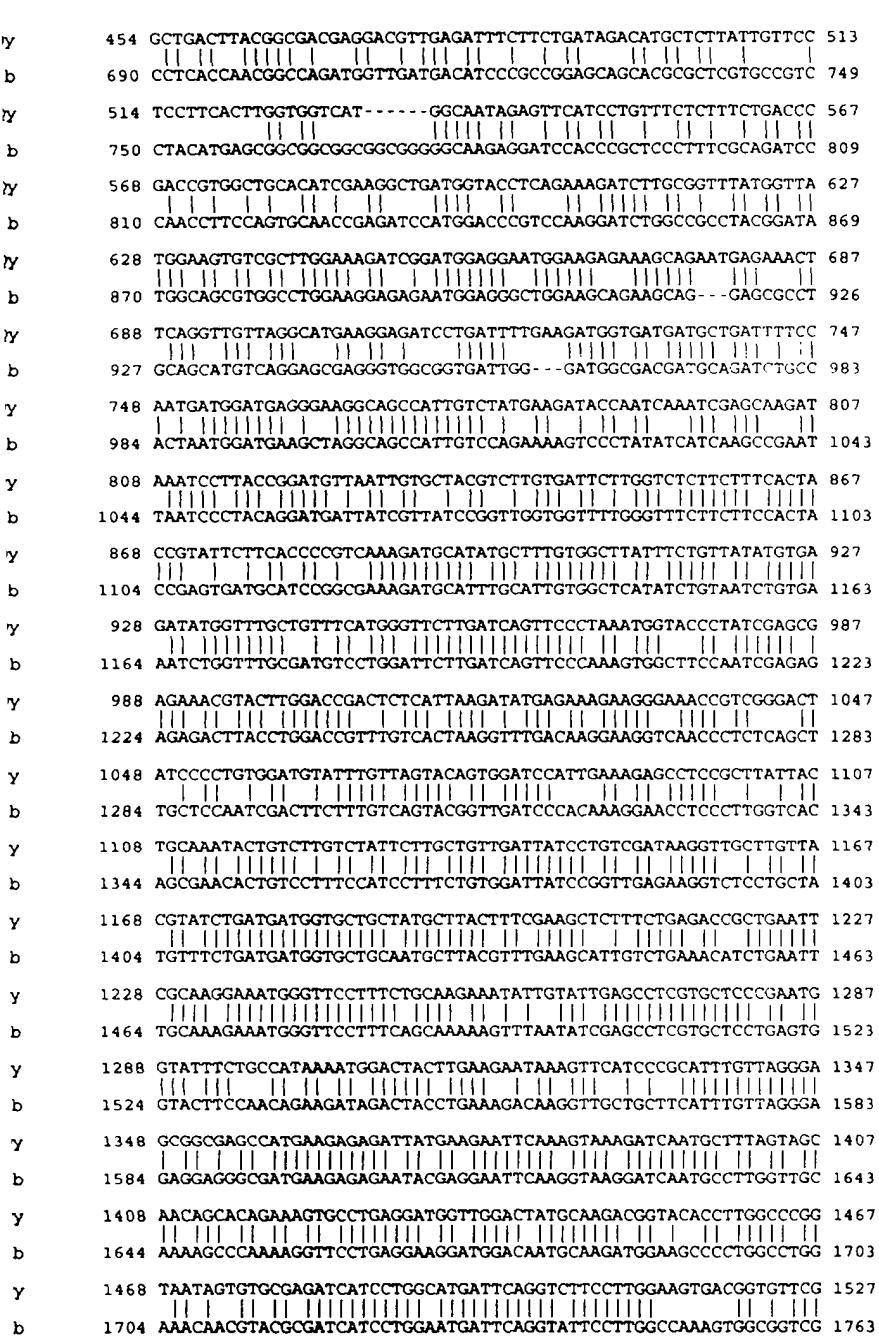
Qy        163 ATGTAACGAATGTGCATTCCCTGTGTGTAGACCTTGCTATGAGTACGAAAGACGAGAAGG 222
              || ||||| || || ||||| || ||  |  | ||||| |||||||  | || || ||
Db        390 GTGCAACGAGTGCGCCTTCCCCGTCTGCCGCGCCTGCTACGAGTACGAGCGCCGCGAGGG 449

Qy        223 CAATCAAGCTTGTCCACAGTGCAAAACCCGTTTCAAACGTCTTAAAGGAAGTCCAAGAGT 282
              |  ||||| || || |||||||  |||||| | ||| || || || ||  | ||  | ||
Db        450 CTCGCAAGCGTGCCCGCAGTGCAGGACCCGCTACAAGCGCCTCAAGGGCTGCCCGCGGGT 509

Qy        283 TGAAGGTGATGAAGAGGAAGATGACATTGATGATTTAGACAATGAGTT---------TGA 333
              |  || || ||||| || | | | || ||  | ||     |||||          ||
Db        510 GGCCGGCGACGAGGAGGAGGACGGCGTCGACGACCTGGAGGGCGAGTTCGGCCTGCAGGA 569

Qy        334 GTATGGAAATAATGGGATTGGATTTGATCAGGTTTCTGAAGGTATGTCAATCTCTCGTCG 393
                  |     | | |   |     |  || | ||    |||   |   | |
Db        570 CGGCGCCGCCCACGAGGACGACCCGCAGTACGTCGCCGAGTCCATGCTCAGGGCGCAGAT 629

Qy        394 CAACTCCGGTTTCCCACAATCTGATTTGGATTCAGCTCCACCTGGCTCTCAGATTCCATT 453
              | || |||   |          |                  || | | |  |  | ||  |
Db        630 GAGCTACGGCCGCGGCGGCGACGCGCACCCCGGCTTCAGCCCCGTCCCCAACGTGCCGCT 689
```


Qy 454 GCTGACTTACGGCGACGAGGACGTTGAGATTTCTTCTGATAGACATGCTCTTATTGTTCC 513
Db 690 CCTCACCAACGGCCAGATGGTTGATGACATCCCGCCGGAGCAGCACGCGCTCGTGCCGTC 749
Qy 514 TCCTTCACTTGGTGGTCAT------GGCAATAGAGTTCATCCTGTTTCTCTTTCTGACCC 567
Db 750 CTACATGAGCGGCGGCGGCGGCGGGGGCAAGAGGATCCACCCGCTCCCTTTCGCAGATCC 809
Qy 568 GACCGTGGCTGCACATCGAAGGCTGATGGTACCTCAGAAAGATCTTGCGGTTTATGGTTA 627
Db 810 CAACCTTCCAGTGCAACCGAGATCCATGGACCCGTCCAAGGATCTGGCCGCCTACGGATA 869
Qy 628 TGGAAGTGTCGCTTGGAAAGATCGGATGGAGGAATGGAAGAGAAAGCAGAATGAGAAACT 687
Db 870 TGGCAGCGTGGCCTGGAAGGAGAGAATGGAGGGCTGGAAGCAGAAGCAG---GAGCGCCT 926
Qy 688 TCAGGTTGTTAGGCATGAAGGAGATCCTGATTTTGAAGATGGTGATGATGCTGATTTTCC 747
Db 927 GCAGCATGTCAGGAGCGAGGGTGGCGGTGATTGG---GATGGCGACGATGCAGATCTGCC 983
Qy 748 AATGATGGATGAGGGAAGGCAGCCATTGTCTATGAAGATACCAATCAAATCGAGCAAGAT 807
Db 984 ACTAATGGATGAAGCTAGGCAGCCATTGTCCAGAAAAGTCCCTATATCATCAAGCCGAAT 1043
Qy 808 AAATCCTTACCGGATGTTAATTGTGCTACGTCTTGTGATTCTTGGTCTCTTCTTTCACTA 867
Db 1044 TAATCCCTACAGGATGATTATCGTTATCCGGTTGGTGGTTTTGGGTTTCTTCTTCCACTA 1103
Qy 868 CCGTATTCTTCACCCCGTCAAAGATGCATATGCTTTGTGGCTTATTTCTGTTATATGTGA 927
Db 1104 CCGAGTGATGCATCCGGCGAAAGATGCATTTGCATTGTGGCTCATATCTGTAATCTGTGA 1163
Qy 928 GATATGGTTTGCTGTTTCATGGGTTCTTGATCAGTTCCCTAAATGGTACCCTATCGAGCG 987
Db 1164 AATCTGGTTTGCGATGTCCTGGATTCTTGATCAGTTCCCAAAGTGGCTTCCAATCGAGAG 1223
Qy 988 AGAAACGTACTTGGACCGACTCTCATTAAGATATGAGAAAGAAGGGAAACCGTCGGGACT 1047
Db 1224 AGAGACTTACCTGGACCGTTTGTCACTAAGGTTTGACAAGGAAGGTCAACCCTCTCAGCT 1283
Qy 1048 ATCCCCTGTGGATGTATTTGTTAGTACAGTGGATCCATTGAAAGAGCCTCCGCTTATTAC 1107
Db 1284 TGCTCCAATCGACTTCTTTGTCAGTACGGTTGATCCCACAAAGGAACCTCCCTTGGTCAC 1343
Qy 1108 TGCAAATACTGTCTTGTCTATTCTTGCTGTTGATTATCCTGTCGATAAGGTTGCTTGTTA 1167
Db 1344 AGCGAACACTGTCCTTTCCATCCTTTCTGTGGATTATCCGGTTGAGAAGGTCTCCTGCTA 1403
Qy 1168 CGTATCTGATGATGGTGCTGCTATGCTTACTTTCGAAGCTCTTTCTGAGACCGCTGAATT 1227
Db 1404 TGTTTCTGATGATGGTGCTGCAATGCTTACGTTTGAAGCATTGTCTGAAACATCTGAATT 1463
Qy 1228 CGCAAGGAAATGGGTTCCTTTCTGCAAGAAATATTGTATTGAGCCTCGTGCTCCCGAATG 1287
Db 1464 TGCAAAGAAATGGGTTCCTTTCAGCAAAAAGTTTAATATCGAGCCTCGTGCTCCTGAGTG 1523
Qy 1288 GTATTTCTGCCATAAAATGGACTACTTGAAGAATAAAGTTCATCCCGCATTTGTTAGGGA 1347
Db 1524 GTACTTCCAACAGAAGATAGACTACCTGAAAGACAAGGTTGCTGCTTCATTTGTTAGGGA 1583
Qy 1348 GCGGCGAGCCATGAAGAGAGATTATGAAGAATTCAAAGTAAAGATCAATGCTTTAGTAGC 1407
Db 1584 GAGGAGGGCGATGAAGAGAGAATACGAGGAATTCAAGGTAAGGATCAATGCCTTGGTTGC 1643
Qy 1408 AACAGCACAGAAAGTGCCTGAGGATGGTTGGACTATGCAAGACGGTACACCTTGGCCCGG 1467
Db 1644 AAAAGCCCAAAAGGTTCCTGAGGAAGGATGGACAATGCAAGATGGAAGCCCCTGGCCTGG 1703
Qy 1468 TAATAGTGTGCGAGATCATCCTGGCATGATTCAGGTCTTCCTTGGAAGTGACGGTGTTCG 1527
Db 1704 AAACAACGTACGCGATCATCCTGGAATGATTCAGGTATTCCTTGGCCAAAGTGGCGGTCG 1763
Qy 1528 TGATGTCGAAAACAACGAGTTGCCTCGATTAGTTTACGTTTCTCGTGAGAAGAGACCCGG 1587
Db 1764 TGATGTGGAAGGAAATGAGTTGCCTCGCCTGGTTTATGTCTCGAGAGAAAAGAGGCCAGG 1823
Qy 1588 ATTTGATCACCATAAGAAGGCTGGAGCTATGAATTCCCTGATACGAGTCTCTGGGGTTCT 1647
Db 1824 TTATAACCATCACAAGAAGGCTGGTGCCATGAATGCACTGGTCCGTGTCTCTGCTGTCTT 1883
Qy 1648 ATCAAATGCTCCTTACCTTCTGAATGTCGATTGTGATCACTACATCAACAATAGCAAAGC 1707
Db 1884 ATCAAATGCTGCATACCTATTGAACTTGGACTGTGATCACTACATCAACAATAGCAAGGC 1943
Qy 1708 TCTTAGAGAAGCAATGTGTTTCATGATGGATCCTCAGTCAGGAAAGAAAATCTGTTATGT 1767
Db 1944 CATAAAAGAGGCTATGTGTTTCATGATGGATCGTTTGGTGGGGAAGAAAGTGTGCTATGT 2003
Qy 1768 TCAGTTCCCTCAAAGGTTCGATGGGATTGATAGGCACGATCGATACTCAAATCGCAATGT 1827
Db 2004 ACAGTTCCCTCAGA[illegible]TGATCGTACTTGCAAAAATGATCGATACGCTAACAGGAACGT 2063
Qy 1828 TGTGTTCTTTGATA[illegible]AATATGAAA[illegible]ATGGGCTACAAGGGCCTATATACGTCGG 1887
Db 2064 TGTCTTTTTTGACATCAACATGAAAGGTTTGGACGGTATTCAAGGACCCATTTATGTGGG 2123
Qy 1888 TACAGGTTGTGTTTTCAGGAGGCAAGCGCTTTACGGATTTGATGCACCGAAGAAGAAGAA 1947
Db 2124 TACTGGATGTGTTTTCAGACGGCAGGCACTGTATGGTTATGATGCTCCTAAAACGAAGAA 2183
Qy 1948 GGGCCCACGTAAGACATGCAATTGCTGGCCAAAATGGTGTCTCCTATGTTTTGGTTCAAG 2007
Db 2184 GCCACCATCAAGAACTTGCAACTGCTGGCCCAAGTGGTGCCTCTCTTGCTGCTGCAGCAG 2243
Qy 2008 AAAGAATCGTAAAGCAAAGACAGTGGCTGCGGATAAGAAGAAGAAGAATAG--------- 2058
Db 2244 GAACAAGAATAAAAAGAAGACTACAAAACCAAAGACGGAGAAGAAGAAAAGATTATTTTT 2303
Qy 2059 ---GGAAGCGTCAAAGCAGATCCACGCATTAGAAAATATCGAAGAGGGCCGCGGTCATAA 2115
Db 2304 CAAGAAAGCAGAAAACCCATCTCCTGCATATGCTTTGGGTGAAATTGATGAAGGTGCTCC 2363
Qy 2116 AGTTCTTAACGTAGAACAGTCAACCGAGGCAATGCAAATGAAGTTGCAGAAGAAATATGG 2175
Db 2364 AGGTGCTGATATCGAGAAGGCCGGAATCGTAAATCAACAGAAACTAGAGAAGAAATTTGG 2423
Qy 2176 GCAGTCTCCTGTATTTGTTGCATCTGCGCGTCTGGAGAATGGTGGGATGGCTAGAAACGC 2235
Db 2424 GCAGTCTTCTGTTTTTGTCGCATCAACACTTCTTGAGAACGGAGGGACCCTGAAGAGCGC 2483
Qy 2236 AAGCCCGGCTTGTCTGCTTAAAGAAGCCATCCAAGTCATTAGTCGCGGATATGAAGATAA 2295
Db 2484 AAGTCCAGCTTCTCTTCTGAAGGAAGCTATACATGTTATCAGCTGCGGCTACGAAGACAA 2543
Qy 2296 AACTGAATGGGGAAAAGAGATTGGGTGGATCTATGGTTCTGTTACCGAAGATATTCTTAC 2355
Db 2544 GACCGACTGGGGAAAAGAGATTGGCTGGATTTACGGATCGATCACAGAGGATATCTTGAC 2603
Qy 2356 GGGTTCTAAGATGCATTCTCATGGTTGGAGACATGTTTATTGTACACCAAAGTTAGCGGC 2415
Db 2604 TGGATTTAAGATGCACTGCCATGGCTGGCGGTCTATTTACTGCATCCCGAAGCGGCCTGC 2663
Qy 2416 TTTCAAAGGATCAGTTCCAATCAATCTTTCGGATCGTCTCCATCAAGTTCTTCGATGGGC 2475
Db 2664 ATTCAAAGGTTCTGCGCCTCTGAACCTTTCCGACCGTCTTCACCAGGTCCTTCGCTGGGC 2723
Qy 2476 GCTTGGGTCGGTTGAGATTTTCTTGAGTAGGCATTGTCCTATTTGGTATGGTTATGGAGG 2535
Db 2724 CCTTGGGTCCGTCGAAATTTTCTTCAGCAAGCACTGCCCACTTTGGTACGGATACGGCGG 2783
Qy 2536 TGGGTTGAAATGGCTTGAGCGGTTGTCCTACATTAACTCTGTGGTTTACCCGTGGACCTC 2595
Db 2784 CGGGCTAAAATTCCTGGAAAGGTTTTCTTATATCAACTCCATCGTTTATCCCTGGACGTC 2843
Qy 2596 TCTACCGCTCATCGTTTACTGTTCTCTCCCTGCCATCTGTCTTCTCACTGGAAAATTCAT 2655

```
Db       2844 CATTCCTCTCCTGGCTTACTGTACCTTGCCTGCCATCTGCCTGCTCACGGGGAAGTTTAT 2903

Qy       2656 CGTTCCCGAGATTAGCAACTATGCGAGTATCCTCTTCATGGCGCTCTTCTCGTCGATTGC 2715
Db       2904 CACACCAGAGCTTACCAATGTCGCCAGTATCTGGTTCATGGCACTTTTCATCTGCATCTC 2963

Qy       2716 AATAACGGGTATTCTCGAGATGCAATGGGGCAAAGTTGGGACCGATGATTGGTGGAGAAA 2775
Db       2964 CGTGACCGGCATCCTGGAAATGAGGTGGAGTGGCGTGGCCATCGACGACTGGTGGAGGAA 3023

Qy       2776 CGAACAGTTTTGGGTCATTGGAGGTGTTTCTGCGCATCTGTTTGCTCTCTTCCAAGGTCT 2835
Db       3024 CGAGCAGTTCTGGGTCATCGGAGGCGTTTCGGCGCATCTGTTCGCGGTGTTCCAGGGCCT 3083

Qy       2836 CCTCAAGGTTCTTGCTGGTGTCGACACTAACTTCACAGTCACATCAAAAGCAGCTGATGA 2895
Db       3084 GCTGAAGGTGTTCGCCGGCATCGACACGAGCTTCACCGTGACGTCGAAGGCCGGGGACGA 3143

Qy       2896 TGGAGAGTTCTCTGACCTTTACCTCTTCAAATGGACTTCACTTCTCATCCCTCCAATGAC 2955
Db       3144 CGAGGAGTTCTCGGAGCTGTACACGTTCAAGTGGACCACCCTGCTGATACCCCCGACCAC 3203

Qy       2956 TCTACTCATCATAAACGTCATTGGAGTCATAGTCGGAGTCTCTGATGCCATCAGCAATGG 3015
Db       3204 GCTCCTCCTGCTGAACTTCATCGGGGTGGTGGCCGGGATCTCGAACGCGATCAACAACGG 3263

Qy       3016 ATACGACTCGTGGGGACCGCTTTTCGGAAGACTGTTCTTTGCACTTTGGGTCATCATTCA 3075
Db       3264 GTACGAGTCGTGGGGCCCCCTGTTCGGGAAGCTCTTCTTCGCCTTCTGGGTGATCGTCCA 3323

Qy       3076 TCTTTACCCGTTCCTTAAAGGTTTGCTTGGGAAACAAGATAGAATGCCAACCATTATTGT 3135
Db       3324 CCTGTACCCGTTCCTCAAGGGTCTGGTGGGGAGGCAGAACAGGACGCCGACGATCGTCAT 3383

Qy       3136 CGTCTGGTCCATCCTCCTGGCCTCGATTCTTACACTTCTTTGGGTCCGGGTTAATCCGTT 3195
Db       3384 CGTCTGGTCCATCCTGCTGGCCTCGATCTTCTCGCTCCTGTGGGTCCGCGTCGACCCGTT 3443

Qy       3196 TGTGGCGAAAGG---CGGTCCTATTCTCGAGATCTGTGGTTTAGACTGCTTGTGATTCGA 3252
Db       3444 CCTCGCCAAGAGCAACGGCCCGCTCCTGGAGGAGTGTGGCCTGGACTGCAACTGAAGTGG 3503

Qy       3253 TTGACCGGTG 3262
Db       3504 GGGCCCCCTG 3513


RESULT 14
AR267557
LOCUS       AR267557               3603 bp    DNA     linear   PAT 10-APR-2003
DEFINITION  Sequence 5 from patent US 6495740.
ACCESSION   AR267557
VERSION     AR267557.1  GI:29697633
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 3603)
  AUTHORS   Arioli,A., Williamson,R.E., Betzner,A.S. and Peng,L.
  TITLE     Manipulation of cellulose and/or .beta.-1,4 Glucan
  JOURNAL   Patent: US 6495740-A 5 17-DEC 2002;
FEATURES             Location/Qualifiers
     source          1. .3603
                     /organism="unknown"
BASE COUNT      950 a    745 c    852 g   1056 t
ORIGIN

  Query Match           37.6%;  Score 1294.8;  DB 6;  Length 3603;
  Best Local Similarity  64.5%;  Pred. No. 0;
  Matches 2074;  Conservative    0;  Mismatches 1082;  Indels   60;  Gaps    7;

Qy          9 GGTTAATCGCCGGTTCTCACAACAGGAATGAGTTTGTCCTCATTAATGCCGATGAGAATG 68

Db         17 GCTTGGTTGCTGGATCCTACCGGAGAAACGAGCTCGTTCGGATCCGACATGAATCTGATG 76

Qy         69 CCCGAATAAGATCAGTCCAAGAGCTGAGTGGACAGACATGTCAAATCTGCAGAGATGAGA 128
Db         77 GCGGGACCAAACCTTTGAAGAATATGAATGGCCAGATATGTCAGATCTGTGGTGATGATG 136

Qy        129 TCGAATTGACTGTTGATGGAGAACCGTTTGTGGCATGTAACGAATGTGCATTCCCTGTGT 188
Db        137 TTGGACTCGCTGAAACTGGAGATGTCTTTGTCGCGTGTAATGAATGTGCCTTCCCTGTGT 196

Qy        189 GTAGACCTTGCTATGAGTACGAAAGACGAGAAGGCAATCAAGCTTGTCCACAGTGCAAAA 248
Db        197 GTCGGCCTTGCTATGAGTACGAGAGGAAAGATGGAACTCAGTGTTGCCCTCAATGCAAGA 256

Qy        249 CCCGTTTCAAACGTCTTAAAGGAAGTCCAAGAGTTGAAGGTGATGAAGAGGAAGATGACA 308
Db        257 CTAGATTCAGACGACACAGGGGGAGTCCTCGTGTTGAAGGAGATGAAGATGAGGATGATG 316

Qy        309 TTGATGATTTAGACAATGAGTTTGAGTATGGAAATAATGGGATTGGATTTGATCAGGTTT 368
Db        317 TTGATGATATCGAGAATGAGTTCAATTA----------CGCCCAGGGAG 355

Qy        369 CTGAAGGTATGTCAATCTCTCGTCGCAACTCCGGTTTCCCACAATCTGATTTGGATTCAG 428
Db        356 CTAACAAGGCGAGACACCAACGCCATGGCGAAGAGTTTTCTTCTTCCTCTAGACATGAAT 415

Qy        429 CTCCACCTGGCTCTCAGATTCCATTGCTGACTTACGGCGACGAGGACGTTGAGATTTCTT 488
Db        416 CTCAACCAATTCCTCTTCTCACCCATGGCCATACGGTTTCTGGAGAGATTCGCACGCCTG 475

Qy        489 CTGATAGACATGCTCTTATTGTTCCTCCTTCACTTGGTGGTCATGGCAATAGAGTTCATC 548
Db        476 ATACACAATCTGTGCGAACTACATCAGGTCCTTTGGGTCCTTCTGACAGGAATGCTATTT 535

Qy        549 CTGTTTCTCTTTCTGACCCGACCGTGGCTGCACATCGAAGGCTGATGGTACCTCAGAAAG 608
Db        536 CATCTCCATATATTGATCCACGGCAACCTGTCCCTGTAAGAATCGTGGACCCGTCAAAAG 595

Qy        609 ATCTTGCGGTTTATGGTTATGGAAGTGTCGCTTGGAAAGATCGGATGGAGGAATGGAAGA 668
Db        596 ACTTGAACTCTTATGGGCTTGGTAATGTTGACTGGAAAGAAAGAGTTGAAGGCTGGAAGC 655

Qy        669 GAAAGCAGAATGAGAAACT----------TCAGGTTGTTAGGCATGAAGGAG 710
Db        656 TGAAGCAGGAGAAAAATATGTTACAGATGACTGGTAAATACCATGAAGGGAAAGGAGGAG 715

Qy        711 ATCCTGATTTTGAAGATGGTGATGATGCTGATTTTCCAATGATGGATGAGGGAAGGCAGC 770
Db        716 AAATTGAAGGGACTGGTTCCAATGGCGAAGAACTCCAAATGGCTGATGATACACGTCTTC 775

Qy        771 CATTGTCTATGAAGATACCAATCAAATCGAGCAAGATAAATCCTTACCGGATGTTAATTG 830
Db        776 CTATGAGTCGTGTGGTGCCTATCCCATCTTCTCGCCTAACCCCTTATCGGGTTGTGATTA 835

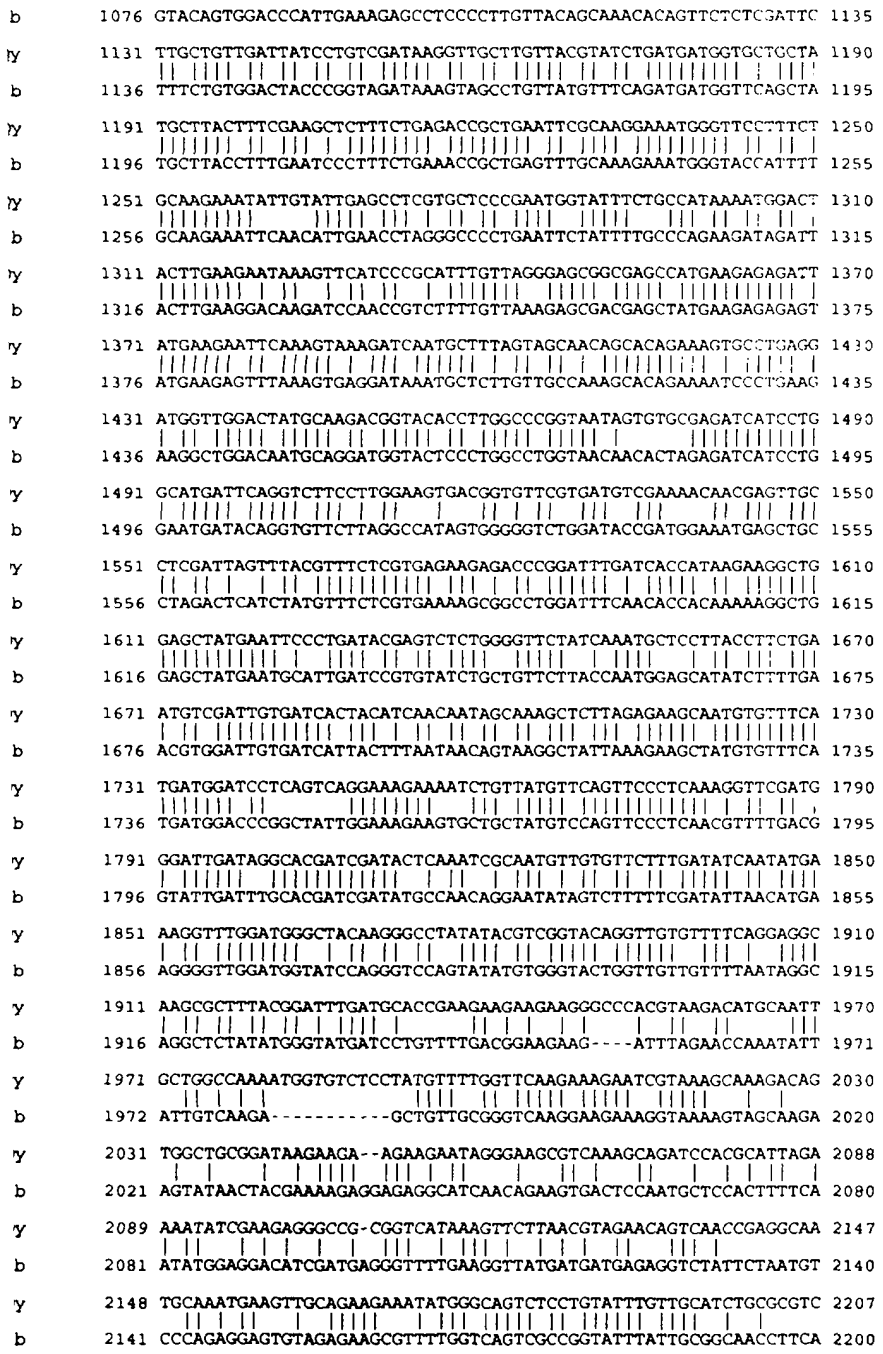
Qy        831 TGCTACGTCTTGTGATTCTTGGTCTCTTCTTTCACTACCGTATTCTTCACCCCGTCAAAG 890
Db        836 TTCTCCGGCTTATCATCTTGTGTTTCTTCTTGCAATATCGTACAACTCACCCTGTGAAAA 895

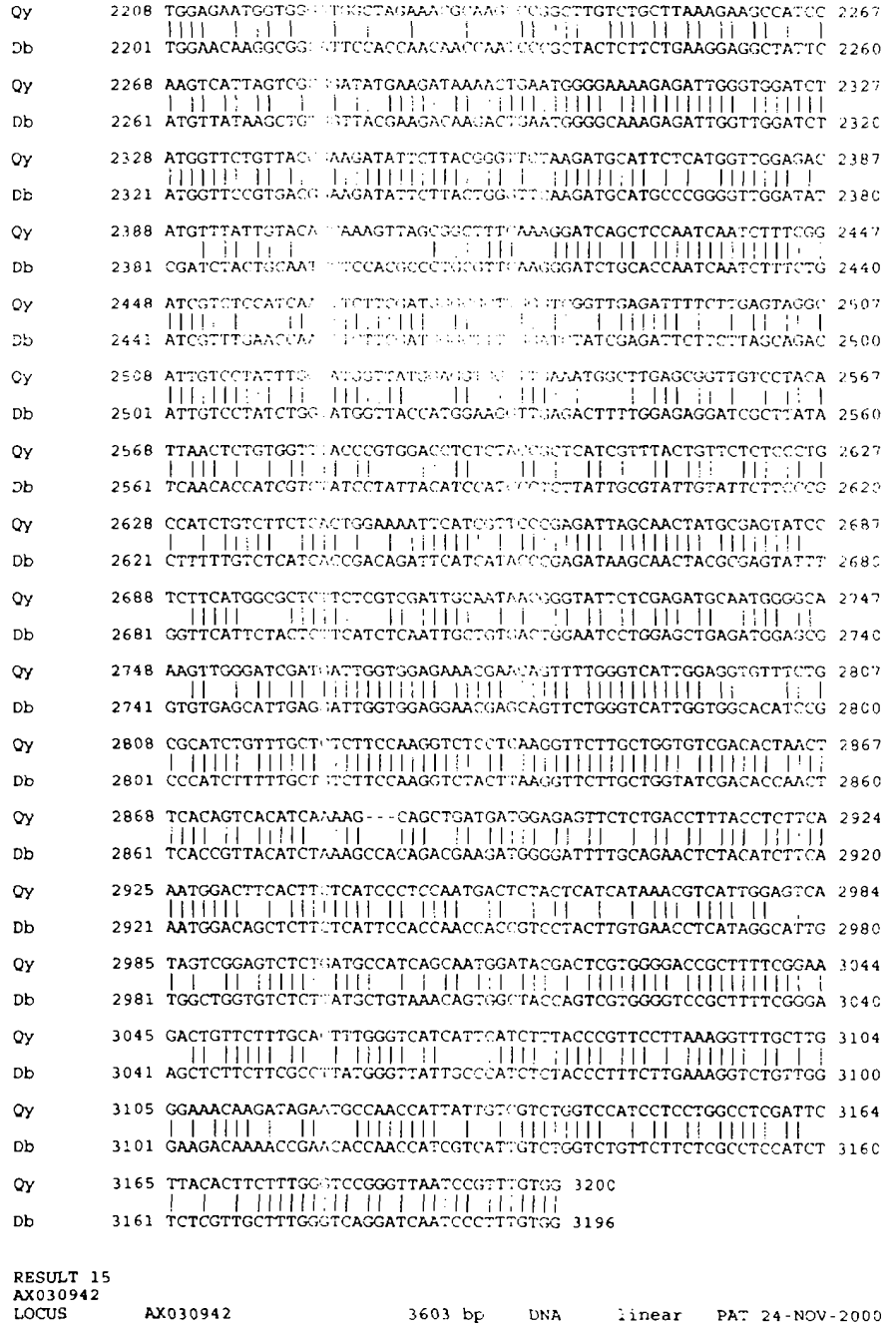
Qy        891 ATGCATATGCTTTGTGGCTTATTTCTGTTATATGTGAGATATGGTTTGCTGTTTCATGGG 950
Db        896 ATGCATATCCTTTGTGGTTGACCTCGGTTATCTGTGAGATCTGGTTTGCATTTTCTTGGC 955

Qy        951 TTCTTGATCAGTTCCCTAAATGGTACCCTATCGAGCGAGAAACGTACTTGGACCGACTCT 1010
Db        956 TTCTTGATCAGTTTCCCAAATGGTACCCCATTAACAGGGAGACTTATCTTGACCGTCTCG 1015

Qy       1011 CATTAAGATATGAGAAAGAAGGGAAACCGTCGGGACTATCCCCTGTGGATGTATTTGTTA 1070
Db       1016 CTATAAGATATGATCGAGACGGTGAACCATCACAGCTCGTTCCTGTTGATGTGTTTGTTA 1075

Qy       1071 GTACAGTGGATCCATTGAAAGAGCCTCCGCTTATTACTGCAAATACTGTCTTGTCTATTC 1130
```





RESULT 15
AX030942
LOCUS AX030942 3603 bp DNA linear PAT 24-NOV-2000

```
DEFINITION  Sequence 5 from Patent WO9800549.
ACCESSION   AX030942 AX030956
VERSION     AX030942.1  GI:10278345
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Williamson,R.E., Peng,L., Arioli,A. and Betzner,A.S.
  TITLE     Manipulation of cellulose and/or beta -1,4-glucan
  JOURNAL   Patent: WO 9800549-A 5 08-JAN-1998;
            WILLIAMSON RICHARD EDWARD (AU) ; PENG LIANGCAI (AU) ; ARIOLI
            ANTONIO (AU) ; UNIV AUSTRALIAN (AU) ; BETZNER ANDREAS STEFAN (AU) ;
            COMMW SCIENT IND RES ORG (AU)
COMMENT     On Oct 15, 2002 this sequence version replaced gi:10278359.
FEATURES             Location/Qualifiers
     source          1..3603
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /strain="COLUMBIA"
                     /db_xref="taxon:3702"
                     /clone="RSW1 cDNA"
     CDS             1..3246
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAC09625.1"
                     /db_xref="GI:10278346"
                     /db_xref="REMTREMBL:CAC09625"
                     /translation="MEASAGLVAGSYRRNELVRIRHESDGGTKPLKNMNGQICQICGD
                     DVGLAETGDVFVACNECAFPVCRPCYEYERKDGTQCCPQCKTRFRRHRGSPRVEGDED
                     EDDVDDIENEFNYAQGANKARHQRHGEEFSSSSRHESQPIPLLTHGHTVSGEIRTPDT
                     QSVRTTSGPLGPSDRNAISSPYIDPRQPVPVRIVDPSKDLNSYGLGNVDWKERVEGWK
                     LKQEKNMLQMTGKYHEGKGGEIEGTGSNGEELQMADDTRLPMSRVVPIPSSRLTPYRV
                     VIILRLIILCFFLQYRTTHPVKNAYPLWLTSVICEIWFAFSWLLDQFPKWYPINRETY
                     LDRLAIRYDRDGEPSQLVPVDVFVSTVDPLKEPPLVTANTVLSILSVDYPVDKVACYV
                     SDDGSAMLTFESLSETAEFAKKWVPFCKKFNIEPRAPEFYFAQKIDYLKDKIQPSFVK
                     ERRAMKREYEEFKVRINALVAKAQKIPEEGWTMQDGTPWPGNNTRDHPGMIQVFLGHS
                     GGLDTDGNELPRLIYVSREKRPGFQHHKKAGAMNALIRVSAVLTNGAYLLNVDCDHYF
                     NNSKAIKEAMCFMMDPATGKKCCYVQFPQRFDGIDLHDRYANRNIVFFDINMKGLDGI
                     QGPVYVGTGCCFNRQALYGYDPVLTEEDLEPNIFVKSCCGSRKKGKSSKKYNYEKRRG
                     INRSDSNAPLFNMEDIDEGFEGYDDERSILMSQRSVEKRFGQSPVFIAATFMEQGGIP
                     PTTNPATLLKEAIHVISCGYEDKTEWGKEIGWIYGSVTEDILTGFKMHARGWISIYCN
                     PPRPAFKGSAPINLSDRLNQVLRWALGSIEILLSRHCPIWYGYHGRLRLLERIAYINT
                     IVYPITSIPLIAYCILPAFCLITDRFIIPEISNYASIWFILLFISIAVTGILELRWSG
                     VSIEDWWRNEQFWVIGGTSAHLFAVFQGLLKVLAGIDTNFTVTSKATDEDGDFAELYI
                     FKWTALLIPPTTVLIVNLIGIVAGVSYAVNSGYQSWGPLFGKLFFALWVIAHLYPFLK
                     GLLGRQNRTPTIVIVWSVLLASIFSLLWVRINPFVAKPNANNFNGKGGVF"
BASE COUNT      950 a    745 c    852 g    [illegible] t
ORIGIN
```

Query Match 37.6%; Score 1294.8; DB 6; Length 3603;
Best Local Similarity 64.5%; Pred. No. 0;
Matches 2074; Conservative 0; Mismatches 1082; Indels 60; Gaps 7;

```
Qy          9 GGTTAATCGCCGGTTCTCACAACAGGAATGAGTTTGTCCTCATTAATGCCGATGAGAATG 68
              | ||  | || ||  ||    | || | | ||  |       ||     |||
Db         17 GCTTGGTTGCTGGATCCTACCGGAGAAACGAGCTCGTTCGCATCCGACATGAATCTGATG 76

Qy         69 CCCGAATAAGATCAGTCCAAGAGCTGAGTGGACAGACATGCAAATCTGCAGAGATGAGA 128
              | | |  | | |  |  |  | ||  | |  |||| |||  | |||||  | |||||
Db         77 GCGGGACCAAACCTTTGAAGAATATGAATGGCCAGATATGTAGATCTGTGGTGATGATG 136

Qy        129 TCGAATTGACTGTTGATGGAGAACCGTTTGTGGCATGTAACGAATGTGCATTCCCTGTGT 188
              | | | |  |||    ||||||     ||||  | |||||  |||||||||
Db        137 TTGGACTCGCTGAAACTGGAGATGTCTTTGTCGCGTGTAATAATGTGCCTTCCCTGTGT 196

Qy        189 GTAGACCTTGCTATGAGTACGAAAGACGAGAAGGCAATCAAGTTGTCCACAGTGCAAAA 248
              || | ||||||||||||||||| ||   |||  | | | |  ||| || |||||
Db        197 GTCGGCCTTGCTATGAGTACGAGAGGAAAGATGGAACTCAAGGTTGCCCTCAATGCAAGA 256

Qy        249 CCCGTTTCAAACGTCTTAAAGGAAGTCAAGAGTTGAAGGCGATGAAGAGGAAGATGACA 308
```

```
Db        257 CTAGATTCAGACGACACAGGGGGAGTCCTCGTGTTGAAGGAGATGAAGATGAGGATGATG 316

Qy        309 TTGATGATTTAGACAATGAGTTTGAGTATGGAAATAATGGGATTGGATTTGATCAGGTTT 368
              |||||||  | ||  ||||||||  | ||                  |  ||||
Db        317 TTGATGATATCGAGAATGAGTTCAATTA--------------------CGCCCAGGGAG 355

Qy        369 CTGAAGGTATGTCAATCTCTCGTCGCAACTCCGGTTTCCCACAATCTGATTTGGATTCAG 428
              ||  |    |  |  |  ||  |   |  |  ||  |   ||  |   ||  |
Db        356 CTAACAAGGCGAGACACCAACGCCATGGCGAAGAGTTTTCTTCTTCCTCTAGACATGAAT 415

Qy        429 CTCCACCTGGCTCTCAGATTCCATTGCTGACTTACGGCGACGAGGACGTTGAGATTTCTT 488
              ||| |||     |||  |  |          |  |     |  ||  ||  |  ||
Db        416 CTCAACCAATTCCTCTTCTCACCCATGGCCATACGGTTTCTGGAGAGATTCGCACGCCTG 475

Qy        489 CTGATAGACATGCTCTTATTGTTCCTCCTTCACTTGGTGGTCATGGCAATAGAGTTCATC 548
              |    |  ||  |  |  |   |  |  | |||  |  || ||  |  |  |
Db        476 ATACACAATCTGTGCGAACTACATCAGGTCCTTTGGGTCCTTCTGACAGGAATGCTATTT 535

Qy        549 CTGTTTCTCTTTCTGACCCGACCGTGGCTGCACATCGAAGGCTGATGGTACCTCAGAAAG 608
              |  | |  | ||| ||        |||  | |  |||  |  |||  ||    ||||
Db        536 CATCTCCATATATTGATCCACGGCAACCTGTCCCTGTAAGAATCGTGGACCCGTCAAAAG 595

Qy        609 ATCTTGCGGTTTATGGTTATGGAAGTGTCGCTTGGAAAGATCGGATGGAGGAATGGAAGA 668
              |  |        |||||||  |||  | |||  |  |||||||  |  | | |  ||||||
Db        596 ACTTGAACTCTTATGGGCTTGGTAATGTTGACTGGAAAGAAAGAGTTGAAGGCTGGAAGC 655

Qy        669 GAAAGCAGAATGAGAAACT----------------TCAGGTTGTTAGGCATGAAGGAG 710
                |||||| |  | ||  |                |  ||  || | | |||||
Db        656 TGAAGCAGGAGAAAAATATGTTACAGATGACTGGTAAATACCATGAAGGGAAAGGAGGAG 715

Qy        711 ATCCTGATTTTGAAGATGGTGATGATGCTGATTTTCCAATGATGGATGAGGGAAGGCAGC 770
              |  |||      | |  |||  |  ||  | | ||||  |||||  | | |  |
Db        716 AAATTGAAGGGACTGGTTCCAATGGCGAAGAACTCCAAATGGCTGATGATACACGTCTTC 775

Qy        771 CATTGTCTATGAAGATACCAATCAAATCGAGCAAGATAAATCCTTACCGGATGTTAATTG 830
              |  ||  |     | | || |||  |||         |||  ||||| ||| |  | |||
Db        776 CTATGAGTCGTGTGGTGCCTATCCCATCTTCTCGCCTAACCCCTTATCGGGTTGTGATTA 835

Qy        831 TGCTACGTCTTGTGATTCTTGGTCTCTTCTTTCACTACCGTATTCTTCACCCCGTCAAAG 890
              | ||  || ||| | ||  |  || |||||| || || ||||    ||||| || |||
Db        836 TTCTCCGGCTTATCATCTTGTGTTTCTTCTTGCAATATCGTACAACTCACCCTGTGAAAA 895

Qy        891 ATGCATATGCTTTGTGGCTTATTTCTGTTATATGTGAGATATGGTTTGCTGTTTCATGGG 950
              |||||||  |||||||  | |  || ||||  |||||||  |||||||  ||||  |||
Db        896 ATGCATATCCTTTGTGGTTGACCTCGGTTATCTGTGAGATCTGGTTTGCATTTTCTTGGC 955

Qy        951 TTCTTGATCAGTTCCCTAAATGGTACCCTATCGAGCGAGAAACGTACTTGGACCGACTCT 1010
              ||||||||||||  || ||||||||||  ||  |  | || || ||  | ||||  |||
Db        956 TTCTTGATCAGTTTCCCAAATGGTACCCCATTAACAGGGAGACTTATCTTGACCGTCTCG 1015

Qy       1011 CATTAAGATATGAGAAAGAAGGGAAACCGTCGGGACTATCCCCTGTGGATGTATTTGTTA 1070
              |  ||||||||||    ||| ||  |||| ||      |||||  |||||  |||||||
Db       1016 CTATAAGATATGATCGAGACGGTGAACCATCACAGCTCGTTCCTGTTGATGTGTTTGTTA 1075

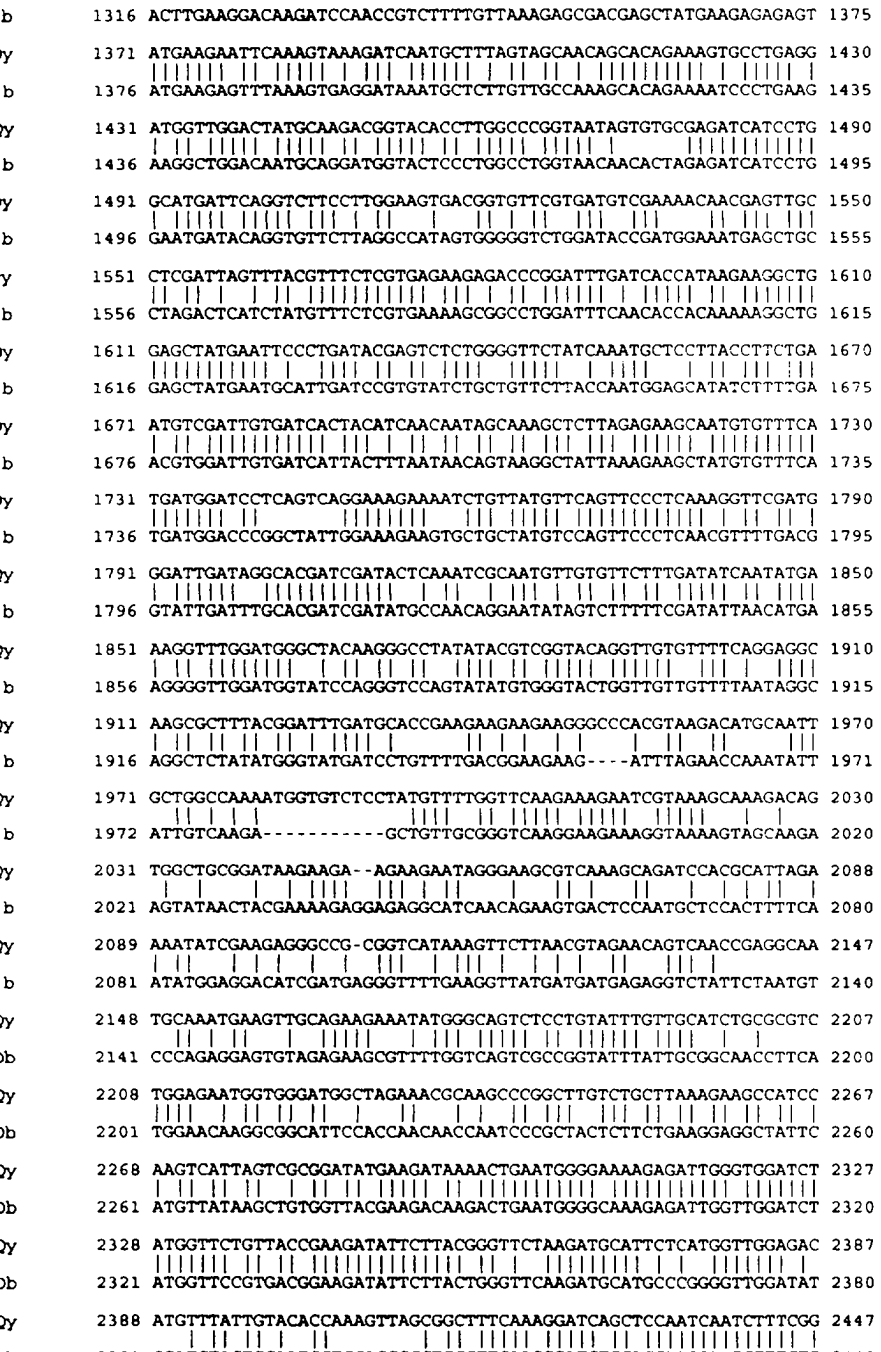
Qy       1071 GTACAGTGGATCCATTGAAAGAGCCTCCGCTTATTACTGCAAATACTGTCTTGTCTATTC 1130
              ||||||||||  |||||||||||||||||  |||  ||||  |||||  || ||  | || ||||
Db       1076 GTACAGTGGACCCATTGAAAGAGCCTCCCCTTGTTACAGCAAACACAGTTCTCTCGATTC 1135

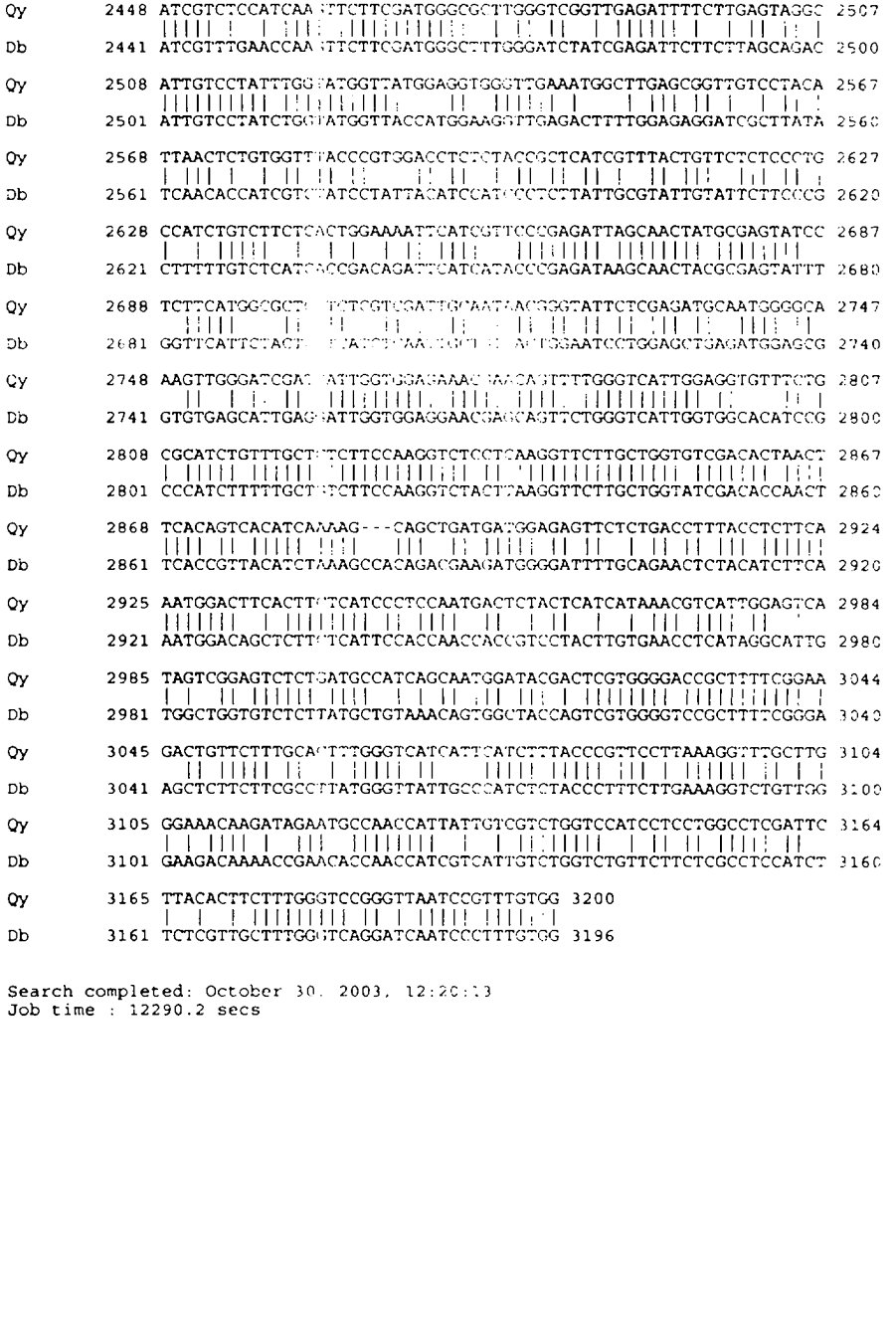
Qy       1131 TTGCTGTTGATTATCCTGTCGATAAGGTTGCTTGTTACGTATCTGATGATGGTGCTGCTA 1190
              ||  |||| || || || || |||||  || || |||||  || || |||||||||  | ||||
Db       1136 TTTCTGTGGACTACCCGGTAGATAAAGTAGCCTGTTATGTTTCAGATGATGGTTCAGCTA 1195

Qy       1191 TGCTTACTTTCGAAGCTCTTTCTGAGACCGCTGAATTCGCAAGGAAATGGGTTCCTTTCT 1250
              |||||||  || |||  | ||||||||  ||||||||  || ||||  |||||||||  || || |
Db       1196 TGCTTACCTTTGAATCCCTTTCTGAAACCGCTGAGTTTGCAAAGAAATGGGTACCATTTT 1255

Qy       1251 GCAAGAAATATTGTATTGAGCCTCGTGCTCCCGAATGGTATTTCTGCCATAAAATGGACT 1310
              |||||||||     ||||| ||| | || || ||||  |||||    ||| || || || |
Db       1256 GCAAGAAATTCAACATTGAACCTAGGGCCCCTGAATTCTATTTTGCCCAGAAGATAGATT 1315

Qy       1311 ACTTGAAGAATAAAGTTCATCCCGCATTTGTTAGGGAGCGGCGAGCCATGAAGAGAGATT 1370
              |||||||  | ||  | ||  ||  | |||||||  |||||  |||||  ||||||||||  |
```





Search completed: October 30, 2003, 12:20:13
Job time : 12290.2 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 30, 2003, 21:53:32 ; Search time 6759.55 Seconds
(without alignments)
12383.170 Million cell updates/sec

Title: US-09-980-043-5
Perfect score: 3444
Sequence: 1 gcggccgcggttaatcgccg.....gtgttgtagatagcggccgc 3444

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 22781392 seqs, 12152238056 residues

Word size : 0

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	593	17.2	593	9	AV782896	AV782896 AV782896
c	2	500	14.5	637	9	AV524115	AV524115 AV524115
c	3	450	13.1	558	9	AV546832	AV546832 AV546832
	4	417	12.1	582	9	AV823575	AV823575 AV823575
c	5	414	12.0	438	9	AB038726	AB038726 AB038726
	6	396	11.5	396	9	AV536569	AV536569 AV536569
c	7	389	11.3	609	9	AV540720	AV540720 AV540720
c	8	378	11.0	430	9	AV784570	AV784570 AV784570
c	9	340	9.9	528	14	W43509	W43509 22886 CD4-1
c	10	322	9.3	542	9	AV564362	AV564362 AV564362
	11	313	9.1	540	14	H36985	H36985 15114 Lambd
	12	303	8.8	303	9	AV529949	AV529949 AV529949
	13	296	8.6	503	14	N96884	N96884 21830 Lambd
c	14	245	7.1	263	9	AV547751	AV547751 AV547751
c	15	238	6.9	380	9	AV546593	AV546593 AV546593
c	16	237	6.9	338	9	AV546302	AV546302 AV546302
c	17	237	6.9	345	9	AV547888	AV547888 AV547888
c	18	213	6.2	317	10	BE038159	BE038159 AA09F07 A
c	19	202	5.9	580	9	AI994360	AI994360 701557762
	20	190	5.5	308	14	Z30729	Z30729 ATTS2475 Ve
	21	187	5.4	510	9	AV527222	AV527222 AV527222
	22	186	5.4	675	9	AV822205	AV822205 AV822205
c	23	183	5.3	404	9	AV808765	AV808765 AV808765
c	24	177	5.1	414	9	AU229295	AU229295 AU229295
c	25	163	4.7	233	9	AV547364	AV547364 AV547364
c	26	161	4.7	416	9	AV808255	AV808255 AV808255
	27	158	4.6	439	9	AV525753	AV525753 AV525753
c	28	157	4.6	395	9	AV539120	AV539120 AV539120
c	29	143	4.2	424	9	AV548715	AV548715 AV548715
c	30	138	4.0	416	9	AI996942	AI996942 701551263
c	31	136	3.9	311	9	AV536189	AV536189 AV536189
c	32	134	3.9	261	9	AV783449	AV783449 AV783449
c	33	119	3.5	351	9	AV548112	AV548112 AV548112
	34	114	3.3	165	28	BH171632	BH171632 SALK_0045
	35	64	1.9	646	9	AU238156	AU238156 AU238156
c	36	53	1.5	690	28	BH999394	BH999394 oei15d11.
c	37	53	1.5	709	28	BZ057757	BZ057757 11f21d11.
c	38	51	1.5	235	14	Z30730	Z30730 ATTS2476 Ve
c	39	47	1.4	716	28	BH967660	BH967660 odj53g10.
c	40	46	1.3	771	29	BZ473322	BZ473322 BONLO91TF
	41	44	1.3	538	14	N96707	N96707 21510 CD4-1
	42	43	1.2	718	28	BH933860	BH933860 odf97c06.
	43	42	1.2	530	13	BQ791285	BQ791285 E4162 Chi
c	44	41	1.2	697	28	BH979329	BH979329 odf95b01.
	45	37	1.1	614	28	BH440864	BH440864 BOGKL87TF

ALIGNMENTS

RESULT 1
AV782896/c
LOCUS AV782896 593 bp mRNA linear EST 28-MAR-2002
DEFINITION AV782896 RAFL5 Arabidopsis thaliana cDNA clone RAFL05-02-P19 3', mRNA sequence.
ACCESSION AV782896
VERSION AV782896.1 GI:19801686
KEYWORDS EST.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids ; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 593)
AUTHORS Seki,M., Narusaka,M., Ishida,J., Kamiya,A., Satou,M., Nakajima,M., Oono,Y., Sakurai,T., Carninci,P., Kawai,J., Itoh,M., Ishii,Y., Arakawa,T., Shibata,K., Shinagawa,A., Muramatsu,M., Hayashizaki,Y. and Shinozaki,K.
TITLE Large scale analysis of Arabidopsis full-length cDNA (2002b)
JOURNAL Unpublished
COMMENT Contact: Motoaki Seki
Plant Functional Genomics Research Group
RIKEN Genomic Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-4359
Fax: 81-298-36-9060
Email: mseki@rtc.riken.go.jp

```
          An Arabidopsis full-length cDNA library was constructed essentially
          as reported previously (Seki et al., 1998). This clone is in a
          modified pBluescript vector as a SstI/XhoI insert. Please visit our
          web site (http://www.gsc.riken.go.jp/e/plant/index _e.html) for
          further details.
EATURES             Location/Qualifiers
     source         1. .593
                    /organism="Arabidopsis thaliana"
                    /mol_type="mRNA"
                    /db_xref="taxon:3702"
                    /clone="RAFL05-02-P19"
                    /dev_stage="rosette plants"
                    /lab_host="SOLR"
                    /clone_lib="RAFL5"
                    /note="Site_1: SstI; Site_2: XhoI; subjected to
                    dehydration-treated(1,2,5,10,24 hr)"
ASE COUNT      188 a    132 c    129 g    144 t
RIGIN

 Query Match             17.2%;  Score 593;  DB 9;  Length 593;
 Best Local Similarity   100.0%;  Pred. No. 3.2e-294;
 Matches  593;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

y       2797 AGGTGTTTCTGCGCATCTGTTTGCTCTCTTCCAAGGTCTCCTCAAGGTTCTTGCTGGTGT 2856
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
b        593 AGGTGTTTCTGCGCATCTGTTTGCTCTCTTCCAAGGTCTCCTCAAGGTTCTTGCTGGTGT 534

y       2857 CGACACTAACTTCACAGTCACATCAAAAGCAGCTGATGATGGAGAGTTCTCTGACCTTTA 2916
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
b        533 CGACACTAACTTCACAGTCACATCAAAAGCAGCTGATGATGGAGAGTTCTCTGACCTTTA 474

y       2917 CCTCTTCAAATGGACTTCACTTCTCATCCCTCCAATGACTCTACTCATCATAAACGTCAT 2976
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
b        473 CCTCTTCAAATGGACTTCACTTCTCATCCCTCCAATGACTCTACTCATCATAAACGTCAT 414

y       2977 TGGAGTCATAGTCGGAGTCTCTGATGCCATCAGCAATGGATACGACTCGTGGGGACCGCT 3036
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
b        413 TGGAGTCATAGTCGGAGTCTCTGATGCCATCAGCAATGGATACGACTCGTGGGGACCGCT 354

y       3037 TTTCGGAAGACTGTTCTTTGCACTTTGGGTCATCATTCATCTTTACCCGTTCCTTAAAGG 3096
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
b        353 TTTCGGAAGACTGTTCTTTGCACTTTGGGTCATCATTCATCTTTACCCGTTCCTTAAAGG 294

y       3097 TTTGCTTGGGAAACAAGATAGAATGCCAACCATTATTGTCGTCTGGTCCATCCTCCTGGC 3156
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
b        293 TTTGCTTGGGAAACAAGATAGAATGCCAACCATTATTGTCGTCTGGTCCATCCTCCTGGC 234

y       3157 CTCGATTCTTACACTTCTTTGGGTCCGGGTTAATCCGTTTGTGGCGAAAGGCGGTCCTAT 3216
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
b        233 CTCGATTCTTACACTTCTTTGGGTCCGGGTTAATCCGTTTGTGGCGAAAGGCGGTCCTAT 174

y       3217 TCTCGAGATCTGTGGTTTAGACTGCTTGTGATTCGATTGACCGGTGGATGGGTTGGTGAA 3276
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
b        173 TCTCGAGATCTGTGGTTTAGACTGCTTGTGATTCGATTGACCGGTGGATGGGTTGGTGAA 114

y       3277 AAAGGTTTAATTCCCACGGATCAAAGAGAGGTAAGAGAGATATTGTTTTACCTCTAAAAG 3336
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
b        113 AAAGGTTTAATTCCCACGGATCAAAGAGAGGTAAGAGAGATATTGTTTTACCTCTAAAAG 54

y       3337 ACTCCTTCATTGTGTTCATTAGATGATGAAAAATGAAAAGAAAAAGAAGATTT 3389
             |||||||||||||||||||||||||||||||||||||||||||||||||||||
b         53 ACTCCTTCATTGTGTTCATTAGATGATGAAAAATGAAAAGAAAAAGAAGATTT 1


ESULT 2
V524115/c
OCUS       AV524115               637 bp    mRNA    linear   EST 01-SEP-2000
EFINITION  AV524115 Arabidopsis thaliana aboveground organs two to six-week
           old Arabidopsis thaliana cDNA clone APZL53c11F 3', mRNA sequence.
CCESSION   AV524115
ERSION     AV524115.1  GI:8683643
EYWORDS    EST.
OURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 637)
  AUTHORS   Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
  TITLE     A large scale analysis of cDNA in Arabidopsis thaliana: Generation
            of 12,028 non-redundant expressed sequence tags from normalized and
            size-selected cDNA libraries
  JOURNAL   DNA Res. 7, 175-180 (2000)
  MEDLINE   20363093
   PUBMED   10907847
COMMENT     Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .637
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /clone="APZL53c11F"
                     /tissue_type="aboveground organs"
                     /dev_stage="two to six-week old"
                     /clone_lib="Arabidopsis thaliana aboveground organs two to
                     six-week old"
                     /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT      221 a    140 c    165 g    111 t
ORIGIN

  Query Match             14.5%;  Score 500;  DB 9;  Length 637;
  Best Local Similarity   100.0%;  Pred. No. 3.5e-246;
  Matches  500;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2937 TTCTCATCCCTCCAATGACTCTACTCATCATAAACGTCATTGGAGTCATAGTCGGAGTCT 2996
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        637 TTCTCATCCCTCCAATGACTCTACTCATCATAAACGTCATTGGAGTCATAGTCGGAGTCT 578

Qy       2997 CTGATGCCATCAGCAATGGATACGACTCGTGGGGACCGCTTTTCGGAAGACTGTTCTTTG 3056
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        577 CTGATGCCATCAGCAATGGATACGACTCGTGGGGACCGCTTTTCGGAAGACTGTTCTTTG 518

Qy       3057 CACTTTGGGTCATCATTCATCTTTACCCGTTCCTTAAAGGTTTGCTTGGGAAACAAGATA 3116
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        517 CACTTTGGGTCATCATTCATCTTTACCCGTTCCTTAAAGGTTTGCTTGGGAAACAAGATA 458

Qy       3117 GAATGCCAACCATTATTGTCGTCTGGTCCATCCTCCTGGCCTCGATTCTTACACTTCTTT 3176
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        457 GAATGCCAACCATTATTGTCGTCTGGTCCATCCTCCTGGCCTCGATTCTTACACTTCTTT 398

Qy       3177 GGGTCCGGGTTAATCCGTTTGTGGCGAAAGGCGGTCCTATTCTCGAGATCTGTGGTTTAG 3236
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        397 GGGTCCGGGTTAATCCGTTTGTGGCGAAAGGCGGTCCTATTCTCGAGATCTGTGGTTTAG 338

Qy       3237 ACTGCTTGTGATTCGATTGACCGGTGGATGGGTTGGTGAAAAAGGTTTAATTCCCACGGA 3296
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        337 ACTGCTTGTGATTCGATTGACCGGTGGATGGGTTGGTGAAAAAGGTTTAATTCCCACGGA 278

Qy       3297 TCAAAGAGAGGTAAGAGAGATATTGTTTTACCTCTAAAAGACTCCTTCATTGTGTTCATT 3356
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        277 TCAAAGAGAGGTAAGAGAGATATTGTTTTACCTCTAAAAGACTCCTTCATTGTGTTCATT 218

Qy       3357 AGATGATGAAAAATGAAAAGAAAAAGAAGATTTAATTTTGTTACGAGAATTGTTATTTTT 3416
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        217 AGATGATGAAAAATGAAAAGAAAAAGAAGATTTAATTTTGTTACGAGAATTGTTATTTTT 158

Qy       3417 GCAAGAATGTGTTGTAGATA 3436
              ||||||||||||||||||||
Db        157 GCAAGAATGTGTTGTAGATA 138
```

RESULT 3
AV546832/c
LOCUS AV546832 558 bp mRNA linear EST 06-SEP-2000
DEFINITION AV546832 Arabidopsis thaliana roots Columbia Arabidopsis thaliana cDNA clone RZL20d11F 3', mRNA sequence.
ACCESSION AV546832
VERSION AV546832.1 GI:8718246
KEYWORDS EST.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids ; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 558)
AUTHORS Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
TITLE A large scale analysis of cDNA in Arabidopsis thaliana: Generation of 12,028 non-redundant expressed sequence tags from normalized and size-selected cDNA libraries
JOURNAL DNA Res. 7, 175-180 (2000)
MEDLINE 20363093
PUBMED 10907847
COMMENT Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES Location/Qualifiers
source 1. .558
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="RZL20d11F"
/tissue_type="roots"
/clone_lib="Arabidopsis thaliana roots Columbia"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT 191 a 131 c 85 g 151 t
ORIGIN

Query Match 13.1%; Score 450; DB 9; Length 558;
Best Local Similarity 100.0%; Pred. No. 2.2e-220;
Matches 450; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       2987 GTCGGAGTCTCTGATGCCATCAGTAATGGATACGACTCGTG-GGACCGCTTTTCGGAAGA 3046
             ||||||||||||||||||   |||| ||| ||||  ||||||||||||||||||
Db        558 GTCGGAGTCTCTGATGCCATCA-AATGGATACGACTCGTCG-GGACCGCTTTTCGGAAGA 499

Qy       3047 CTGTTCTTTGCACTTTGGGTCA-ATTTACTTTTACCCGTC-TTAAAGGTTTGCTTGGG 3106
             ||||||||||||||||||||||  |||| |||||||||| ||||||||||||||||||
Db        498 CTGTTCTTTGCACTTTGGGTCATCATTCATCTTTACCCGTC-TTAAAGGTTTGCTTGGG 439

Qy       3107 AAACAAGATAGAATGCCAACCATTATTGTCGTCTGGTCCAT-CTCCTGGCCTCGATTCTT 3166
             ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db        438 AAACAAGATAGAATGCCAACCATTATTGTCGTCTGGTCCAT-CTCCTGGCCTCGATTCTT 379

Qy       3167 ACACTTCTTTGGGTCCGGGTTAATCCGTTTGTGGCGAAAGG-GGTCCTATTCTCGAGATC 3226
             |||||||||||||||||||| |||||||||||||||||||| ||||||||||||||||||
Db        378 ACACTTCTTTGGGTCCGGGTTAATCCGTTTGTGGCGAAAGG-GGTCCTATTCTCGAGATC 319

Qy       3227 TGTGGTTTAGACTGCTTGTGATTCGATTGACCGGTGGATGG-TTGGTGAAAAAGGTTTAA 3286
             |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db        318 TGTGGTTTAGACTGCTTGTGATTCGATTGACCGGTGGATGG-TTGGTGAAAAAGGTTTAA 259

Qy       3287 TTCCCACGGATCAAAGAGAGGTAAGAGAGATATTGTTTTAC-TCTAAAAGACTCCTTCAT 3346
             |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db        258 TTCCCACGGATCAAAGAGAGGTAAGAGAGATATTGTTTTAC-TCTAAAAGACTCCTTCAT 199

Qy       3347 TGTGTTCATTAGATGATGAAAAATGAAAAGAAAAAGAAGAT-AATTTTGTTACGAGAAT 3406
             |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db        198 TGTGTTCATTAGATGATGAAAAATGAAAAGAAAAAGAAGAT-AATTTTGTTACGAGAAT 139
```

Qy 3407 TGTTATTTTTGCAAGAATGTGTTGTAGATA 3436
||||||||||||||||||||||||||||||
Db 138 TGTTATTTTTGCAAGAATGTGTTGTAGATA 109

RESULT 4
AV823575
LOCUS AV823575 582 bp mRNA linear EST 01-APR-2002
DEFINITION AV823575 RAFL5 Arabidopsis thaliana cDNA clone RAFL05-21-E11 5', mRNA sequence.
ACCESSION AV823575
VERSION AV823575.1 GI:19865635
KEYWORDS EST.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids ; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 582)
AUTHORS Seki,M., Narusaka,M., Ishida,J., Kamiya,A., Satou,M., Nakajima,M., Oono,Y., Sakurai,T., Carninci,P., Kawai,J., Itoh,M., Ishii,Y., Arakawa,T., Shibata,K., Shinagawa,A., Muramatsu,M., Hayashizaki,Y. and Shinozaki,K.
TITLE Large scale analysis of Arabidopsis full-length cDNA (2002b)
JOURNAL Unpublished
COMMENT Contact: Motoaki Seki
Plant Functional Genomics Research Group
RIKEN Genomic Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-4359
Fax: 81-298-36-9060
Email: mseki@rtc.riken.go.jp
An Arabidopsis full-length cDNA library was constructed essentially as reported previously (Seki et al., 1998).This clone is in a modified pBluescript vector as a SstI/XhoI insert. Please visit our web site (http://www.gsc.riken.go.jp/e/plant/index _e.html) for further details.
FEATURES Location/Qualifiers
source 1. .582
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/clone="RAFL05-21-E11"
/dev_stage="rosette plants"
/lab_host="SOLR"
/clone_lib="RAFL5"
/note="Site_1: SstI; Site_2: XhoI; subjected to dehydration-treated(1,2,5,10,24 hr)"
BASE COUNT 151 a 114 c 143 g 171 t 3 others
ORIGIN

Query Match 12.1%; Score 417; DB 9; Length 582;
Best Local Similarity 100.0%; Pred. No. 2.5e-203;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1118 GTCTTGTCTATTCTTGCTGTTGATTATCCTGTCGATAAGGTTGCTTGTTACGTATCTGAT 1177
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         94 GTCTTGTCTATTCTTGCTGTTGATTATCCTGTCGATAAGGTTGCTTGTTACGTATCTGAT 153

Qy       1178 GATGGTGCTGCTATGCTTACTTTCGAAGCTCTTTCTGAGACCGCTGAATTCGCAAGGAAA 1237
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        154 GATGGTGCTGCTATGCTTACTTTCGAAGCTCTTTCTGAGACCGCTGAATTCGCAAGGAAA 213

Qy       1238 TGGGTTCCTTTCTGCAAGAAATATTGTATTGAGCCTCGTGCTCCCGAATGGTATTTCTGC 1297
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        214 TGGGTTCCTTTCTGCAAGAAATATTGTATTGAGCCTCGTGCTCCCGAATGGTATTTCTGC 273

Qy       1298 CATAAAATGGACTACTTGAAGAATAAAGTTCATCCCGCATTTGTTAGGGAGCGGCGAGCC 1357
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        274 CATAAAATGGACTACTTGAAGAATAAAGTTCATCCCGCATTTGTTAGGGAGCGGCGAGCC 333

Qy       1358 ATGAAGAGAGATTATGAAGAATTCAAAGTAAAGATCAATGCTTTAGTAGCAACAGCACAG 1417
             ||||||||||:||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
b        334 ATGAAGAGAGATTATGAAGAATTCAAAGTAAAGATCAATGCTTTAGTAGCAACAGCACAG 393
y       1418 AAAGTGCCTGAGGATGGTTGGACTATGCAAGACGGTACACCTTGGCCCGGTAATAGTGTG 1477
b        394 AAAGTGCCTGAGGATGGTTGGACTATGCAAGACGGTACACCTTGGCCCGGTAATAGTGTG 453
y       1478 CGAGATCATCCTGGCATGATTCAGGTCTTCCTTGGAAGTGACGGTGTTCGTGATGTC 1534
b        454 CGAGATCATCCTGGCATGATTCAGGTCTTCCTTGGAAGTGACGGTGTTCGTGATGTC 510

ESULT 5
B038726/c
OCUS       AB038726                438 bp    mRNA    linear   EST 14-NOV-2000
EFINITION  AB038726 Arabidopsis thaliana Above-ground organ from two to
           six-week old plants Columbia Arabidopsis thaliana cDNA clone
           APZ31h04_f 3', mRNA sequence.
CCESSION   AB038726
ERSION     AB038726.1  GI:7212553
EYWORDS    EST.
OURCE      Arabidopsis thaliana (thale cress)
 ORGANISM  Arabidopsis thaliana
           Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
           Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
           ; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
EFERENCE   1  (bases 1 to 438)
 AUTHORS   Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
 TITLE     A large scale analysis of cDNA in Arabidopsis thaliana: Generation
           of 12,028 non-redundant expressed sequence tags from normalized and
           size-selected cDNA libraries
 JOURNAL   DNA Res. 7, 175-180 (2000)
 MEDLINE   20363093
  PUBMED   10907847
OMMENT     Contact: Erika Asamizu
           The First Laboratory for Plant Gene Research
           Kazusa DNA Research Institute
           Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
           Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
EATURES             Location/Qualifiers
     source         1. .438
                    /organism="Arabidopsis thaliana"
                    /mol_type="mRNA"
                    /strain="Columbia"
                    /db_xref="taxon:3702"
                    /clone="APZ31h04_f"
                    /tissue_type="aboveground organ"
                    /dev_stage="two to six-week old plants"
                    /clone_lib="Arabidopsis thaliana Above-ground organ from
                    two to six-week old plants Columbia"
                    /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                    XhoI"
ASE COUNT      144 a    105 c     77 g    112 t
RIGIN

  Query Match             12.0%;  Score 414;  DB 9;  Length 438;
  Best Local Similarity   100.0%;  Pred. No. 7.8e-202;
  Matches  414;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

)y      3006 TCAGCAATGGATACGACTCGTGGGGACCGCTTTTCGGAAGACTGTTCTTTGCACTTTGGG 3065
)b       438 TCAGCAATGGATACGACTCGTGGGGACCGCTTTTCGGAAGACTGTTCTTTGCACTTTGGG 379
)y      3066 TCATCATTCATCTTTACCCGTTCCTTAAAGGTTTGCTTGGGAAACAAGATAGAATGCCAA 3125
)b       378 TCATCATTCATCTTTACCCGTTCCTTAAAGGTTTGCTTGGGAAACAAGATAGAATGCCAA 319
)y      3126 CCATTATTGTCGTCTGGTCCATCCTCCTGGCCTCGATTCTTACACTTCTTTGGGTCCGGG 3185
)b       318 CCATTATTGTCGTCTGGTCCATCCTCCTGGCCTCGATTCTTACACTTCTTTGGGTCCGGG 259
)y      3186 TTAATCCGTTTGTGGCGAAAGGCGGTCCTATTCTCGAGATCTGTGGTTTAGACTGCTTGT 3245
)b       258 TTAATCCGTTTGTGGCGAAAGGCGGTCCTATTCTCGAGATCTGTGGTTTAGACTGCTTGT 199

Qy      3246 GATTCGATTGACC[illegible]GGATGGGTTGGTGAAAAAGGTTTAATTCCCACGGATCAAAGAGA 3305
Db       198 GATTCGATTGACC[illegible]GGATGGGTTGGTGAAAAAGGTTTAATTCCCACGGATCAAAGAGA 139

Qy      3306 GGTAAGAGAGATAT[illegible]GTTTTACCTCTAAAAGACTCCTTCATTGTGTTCATTAGATGATGA 3365
Db       138 GGTAAGAGAGATAT[illegible]GTTTTACCTCTAAAAGACTCCTTCATTGTGTTCATTAGATGATGA 79

Qy      3366 AAAATGAAAAGAAA[illegible]AGAAGATTTAATTTTGTTACGAGAATTGTTATTTTTGCA 3419
Db         78 AAAATGAAAAGAAA[illegible]AGAAGATTTAATTTTGTTACGAGAATTGTTATTTTTGCA 25

RESULT 6
AV536569
LOCUS       AV536569                396 bp    mRNA    linear   EST 06-SEP-2000
DEFINITION  AV536569 Arabidopsis thaliana liquid cultured seedlings Columbia
            Arabidopsis tha[illegible]ana cDNA clone pAZNII0292R 5', mRNA sequence
ACCESSION   AV536569
VERSION     AV536569.1  GI:[illegible]
KEYWORDS    EST.
SOURCE      Arabidopsis tha[illegible]ana (thale cress)
  ORGANISM  Arabidopsis tha[illegible]ana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 396)
  AUTHORS   Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
  TITLE     A large scale analysis of cDNA in Arabidopsis thaliana: Generation
            of 12,028 non-redundant expressed sequence tags from normalized and
            size-selected cDNA libraries
  JOURNAL   DNA Res. 7, 175-180 (2000)
  MEDLINE   20363093
   PUBMED   10907847
COMMENT     Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .396
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /clone="pAZNII0292R"
                     /tissue_type="liquid-cultured seedlings"
                     /clone_lib="Arabidopsis thaliana liquid-cultured seedlings
                     Columbia"
                     /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT       82 a     94 c     95 g    125 t
ORIGIN

  Query Match             12.5%;  Score 396;  DB 9;  Length 396;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-192;
  Matches  396;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2531 GGAGGTGGGTTGAAATGGCTTGAGCGGTTGTCCTACATTAACTCTGTGGTTTACCCGTGG 2590
Db         1 GGAGGTGGGTTGAAATGGCTTGAGCGGTTGTCCTACATTAACTCTGTGGTTTACCCGTGG 60

Qy      2591 ACCTCTCTACCGCT[illegible]ATCGTTTACTGTTCTCTCCGTGCCATCTGTCTTCTCACTGGAAAA 2650
Db         61 ACCTCTCTACCGCT[illegible]ATCGTTTACTGTTCTCTCCGTGCCATCTGTCTTCTCACTGGAAAA 120

Qy      2651 TTCATCGTTCCCGAGATTAGCAACTATGCGAGTATCCTCTTCATGGCGCTCTTCTCGTCG 2710
Db       121 TTCATCGTTCCCGAGATTAGCAACTATGCGAGTATCCTCTTCATGGCGCTCTTCTCGTCG 180

Qy      2711 ATTGCAATAACGGGTATTCTCGAGATGCAATGGGGCAAAGTTGGGATCGATGATTGGTGG 2770
```

```
Db        181 ATTGCAATAACGGGTATTCTCGAGATGCAATGGGGCAAAGTTGGGATCGATGATTGGTGG 240

Qy       2771 AGAAACGAACAGTTTTGGGTCATTGGAGGTGTTTCTGCGCATCTGTTTGCTCTCTTCCAA 2830
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGAAACGAACAGTTTTGGGTCATTGGAGGTGTTTCTGCGCATCTGTTTGCTCTCTTCCAA 300

Qy       2831 GGTCTCCTCAAGGTTCTTGCTGGTGTCGACACTAACTTCACAGTCACATCAAAAGCAGCT 2890
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GGTCTCCTCAAGGTTCTTGCTGGTGTCGACACTAACTTCACAGTCACATCAAAAGCAGCT 360

Qy       2891 GATGATGGAGAGTTCTCTGACCTTTACCTCTTCAAA 2926
              ||||||||||||||||||||||||||||||||||||
Db        361 GATGATGGAGAGTTCTCTGACCTTTACCTCTTCAAA 396


RESULT 7
AV540720/c
LOCUS       AV540720                609 bp    mRNA    linear   EST 07-SEP-2000
DEFINITION  AV540720 Arabidopsis thaliana roots Columbia Arabidopsis thaliana
            cDNA clone RZ154b07F 3', mRNA sequence.
ACCESSION   AV540720
VERSION     AV540720.1  GI:8702478
KEYWORDS    EST.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 609)
  AUTHORS   Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
  TITLE     A large scale analysis of cDNA in Arabidopsis thaliana: Generation
            of 12,028 non-redundant expressed sequence tags from normalized and
            size-selected cDNA libraries
  JOURNAL   DNA Res. 7, 175-180 (2000)
  MEDLINE   20363093
   PUBMED   10907847
COMMENT     Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .609
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /clone="RZ154b07F"
                     /tissue_type="roots"
                     /clone_lib="Arabidopsis thaliana roots Columbia"
                     /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT      219 a    135 c     90 g    165 t
ORIGIN

  Query Match           11.3%;  Score 389;  DB 9;  Length 609;
  Best Local Similarity   99.8%;  Pred. No. 7.2e-189;
  Matches  439;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2997 CTGATGCCATCAGCAATGGATACGACTCGTGGGGACCGCTTTTCGGAAGACTGTTCTTTG 3056
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        609 CTGATGCCATCAGCAATGGATACGACTCGTGGGGACCGCTTTTCGGAAGACTGTTCTTTG 550

Qy       3057 CACTTTGGGTCATCATTCATCTTTACCCGTTCCTTAAAGGTTTGCTTGGGAAACAAGATA 3116
              |||||||||||||||||||||||| |||| |||||||| ||||||||||||||||||||
Db        549 CACTTTGGGTCATCATTCATCTTTAGCCGTTCCTTAAAGGTTTGCTTGGGAAACAAGATA 490

Qy       3117 GAATGCCAACCATTATTGTCGTCTGGTCCATCCTCCTGGCCTCGATTCTTACACTTCTTT 3176
              |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db        489 GAATGCCAACCATTATTGTCGTCTGGTCCATCCTCCTGGCCTCGATTCTTACACTTCTTT 430

Qy       3177 GGGTCCGGGTTAATCCGTTTGTGGCGAAAGGCGGTCCTATTCTCGAGATCTGTGGTTTAG 3236
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        429 GGGTCCGGGTTAATCCGTTTGTGGCGAAAGGCGGTCCTATTCTCGAGATCTGTGGTTTAG 370

Qy       3237 ACTGCTTGTGATTCGATTGACCGGTGGATGGGTTGGTGAAAAAGGTTTAATTCCCACGGA 3296
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        369 ACTGCTTGTGATTCGATTGACCGGTGGATGGGTTGGTGAAAAAGGTTTAATTCCCACGGA 310

Qy       3297 TCAAAGAGAGGTAAGAGAGATATTGTTTTACCTCTAAAAGACTCCTTCATTGTGTTCATT 3356
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        309 TCAAAGAGAGGTAAGAGAGATATTGTTTTACCTCTAAAAGACTCCTTCATTGTGTTCATT 250

Qy       3357 AGATGATGAAAAATGAAAAGAAAAAGAAGATTTAATTTTGTTACGAGAATTGTTATTTTT 3416
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        249 AGATGATGAAAAATGAAAAGAAAAAGAAGATTTAATTTTGTTACGAGAATTGTTATTTTT 190

Qy       3417 GCAAGAATGTGTTGTAGATA 3436
              ||||||||||||||||||||
Db        189 GCAAGAATGTGTTGTAGATA 170


RESULT 8
AV784570/c
LOCUS       AV784570                430 bp    mRNA    linear   EST 28-MAR-2002
DEFINITION  AV784570 RAFL5 Arabidopsis thaliana cDNA clone RAFL05-21-E11 3',
            mRNA sequence.
ACCESSION   AV784570
VERSION     AV784570.1  GI:19803360
KEYWORDS    EST.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 430)
  AUTHORS   Seki,M., Narusaka,M., Ishida,J., Kamiya,A., Satou,M., Nakajima,M.,
            Oono,Y., Sakurai,T., Carninci,P., Kawai,J., Itoh,M., Ishii,Y.,
            Arakawa,T., Shibata,K., Shinagawa,A., Muramatsu,M., Hayashizaki,Y.
            and Shinozaki,K.
  TITLE     Large scale analysis of Arabidopsis full-length cDNA (2002b)
  JOURNAL   Unpublished
COMMENT     Contact: Motoaki Seki
            Plant Functional Genomics Research Group
            RIKEN Genomic Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: 81-298-36-4359
            Fax: 81-298-36-9060
            Email: mseki@rtc.riken.go.jp
            An Arabidopsis full-length cDNA library was constructed essentially
            as reported previously (Seki et al., 1998). This clone is in a
            modified pBluescript vector as a SstI/XhoI insert. Please visit our
            web site (http://www.gsc.riken.go.jp/e/plant/index _e.html) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .430
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
                     /clone="RAFL05-21-E11"
                     /dev_stage="rosette plants"
                     /lab_host="SOLR"
                     /clone_lib="RAFL5"
                     /note="Site_1: SstI; Site_2: XhoI; subjected to
                     dehydration-treated(1,2,5,10,24 hr)"
BASE COUNT      139 a    102 c     93 g     96 t
ORIGIN

  Query Match           11.0%;  Score 378;  DB 9;  Length 430;
  Best Local Similarity   99.8%;  Pred. No. 3e-183;
  Matches  428;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2936 CTTCTCATCCCTCCAATGACTCTACTCATCATAAACGTCATTGGAGTCATAGTCGGAGTC 2995
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        430 CTTCTCATCCCTCCAATGACTCTACTCATCATAAACGTCATTGGAGTCATAGTCGGAGTC 371
```

```
Qy       2996 TCTGATGCCATCAGCAATGGATACGACTCGTGGGGACCGCTTTTCGGAAGACTGTTCTTT 3055
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        370 TCTGATGCCATCAGCAATGGATACGACTCGTGGGGACCGCTTTTCGGAAGACTGTTCTTT 311

Qy       3056 GCACTTTGGGTCATCATTCATCTTTACCCGTTCCTTAAAGGTTTGCTTGGGAAACAAGAT 3115
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        310 GCACTTTGGGTCATCATTCATCTTTACCCGTTCCTTAAAGGTTTGCTTGGGAAACAAGAT 251

Qy       3116 AGAATGCCAACCATTATTGTCGTCTGGTCCATCCTCCTGGCCTCGATTCTTACACTTCTT 3175
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        250 AGAATGCCAACCATTATTGTCGTCTGGTCCATCCTCCTGGCCTCGATTCTTACACTTCTT 191

Qy       3176 TGGGTCCGGGTTAATCCGTTTGTGGCGAAAGGCGGTCCTATTCTCGAGATCTGTGGTTTA 3235
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        190 TGGGTCCGGGTTAATCCGTTTGTGGCGAAAGGCGGTCCTATTCTCGAGATCTGTGGTTTA 131

Qy       3236 GACTGCTTGTGATTCGATTGACCGGTGGATGGGTTGGTGAAAAAGGTTTAATTCCCACGG 3295
             |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db        130 GACTGCTTGTGATTCGATTGACCGGTGGATGGGTTGGTGAGAAAGGTTTAATTCCCACGG 71

Qy       3296 ATCAAAGAGAGGTAAGAGAGATATTGTTTTACCTCTAAAAGACTCCTTCATTGTGTTCAT 3355
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         70 ATCAAAGAGAGGTAAGAGAGATATTGTTTTACCTCTAAAAGACTCCTTCATTGTGTTCAT 11

Qy       3356 TAGATGATG 3364
             |||||||||
Db         10 TAGATGATG 2
```

```
ESULT 9
43509/c
OCUS        W43509                  528 bp    mRNA    linear   EST 05-JAN-1998
EFINITION   22886 CD4-15 Arabidopsis thaliana cDNA clone G12G11T7, mRNA
            sequence.
CCESSION    W43509
ERSION      W43509.1  GI:2748826
EYWORDS     EST.
OURCE       Arabidopsis thaliana (thale cress)
 ORGANISM   Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
EFERENCE    1  (bases 1 to 528)
 AUTHORS    Newman,T., deBruijn,F.J., Green,P., Keegstra,K., Kende,H., McIntosh
            ,L., Ohlrogge,J., Raikhel,N., Somerville,S., Thomashow,M., Retzel
            ,E. and Somerville,C.
 TITLE      Genes galore: a summary of methods for accessing results from
            large-scale partial sequencing of anonymous Arabidopsis cDNA clones
 JOURNAL    Plant Physiol. 106, 1241-1255 (1994)
 MEDLINE    95148729
  PUBMED    7846151
OMMENT      On Jan 5, 1998 this sequence version replaced gi:1327977.
            Contact: Thomas Newman
            MSU-DOE Plant Research Laboratory
            Michigan State University
            MSU-DOE-PRL, Michigan State University,Plant Biology Bldg.,E.
            Lansing,Mi
            Tel: 517-353-0854
            Fax: 517-353-9168
            Email: 22313tcn@ibm.cl.msu.edu
            Seq primer: T7.
EATURES             Location/Qualifiers
     source         1. .528
                    /organism="Arabidopsis thaliana"
                    /mol_type="mRNA"
                    /strain="Columbia"
                    /db_xref="taxon:3702"
                    /clone="G12G11T7"
                    /tissue_type="seedling hypocotyl"
                    /dev_stage="3 day-old"
                    /clone_lib="CD4-15"
                    /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                    EcoRI; Using 5 ug of polyadenylated mRNA from 3 day-old
                    Arabidopsis thaliana (Columbia) seedling hypocotyls as
                    template and oligo d(T) as primer, first strand synthesis
                    was catalyzed by Moloney murine leukemia virus reverse
                    transcriptase (Pharmacia).  Second-strand cDNA was made
                    using the procedure of Gubler and Hoffman (1983) except
                    that DNA ligase was omitted.  After the second strand
                    reaction, the ends of the cDNA were made blunt with Klenow
                    fragment and EcoRI/NotI adapters (Pharmacia) were ligated
                    to each end.  The cDNA was purified from unligated
                    adapters by spun-column chromatography using sephacryl
                    s-300 and size-fractionated on a 1% low melting point
                    mini-gel.  Size selected cDNAs (2 - 3 kb) were removed
                    from the gel using agarase (New England Biolabs),
                    phenol:choloroform extracted and precipitated using 0.3 M
                    NaOAc (pH 7)/ethanol.  A portion of each cDNA
                    size-fraction (0.1 ug) was co-precipitated with 1 ug of
                    lambdaZapII (Stratagene EcoRI digested, dephosphorylated
                    arms and then ligated in a volume of 4 ul overnight. Each
                    ligation mix was packaged in vitro using Gigapack II gold
                    packaging extract (Stratagene).  We have determined that
                    although first strand cDNA synthesis was initiated using
                    dT, almost all of the cDNAs begin 8-10 bp from the poly-A
                    tail.  The reason for the loss of the poly-A tail is most
                    likely due to lower than anticipated nucleotide levels
                    during the Klenow repair of ragged ends before the
                    addition of linkers (3'-5' exo instead of 5'-3' pol).
                    When this library is used please reference the ABRC and:
                    Kieber, J. et al. (1993) Cell 72:427-441."
BASE COUNT      154 a    116 c    109 g    136 t     13 others
ORIGIN
```

```
  Query Match             9.9%;  Score 340;  DB 14;  Length 528;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-163;
  Matches  340;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3097 TTTGCTTGGGAAACAAGATAGAATGCCAACCATTATTGTCGTCTGGTCCATCCTCCTGGC 3156
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        357 TTTGCTTGGGAAACAAGATAGAATGCCAACCATTATTGTCGTCTGGTCCATCCTCCTGGC 298

Qy       3157 CTCGATTCTTACACTTCTTTGGGTCCGGGTTAATCCGTTTGTGGCGAAAGGCGGTCCTAT 3216
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        297 CTCGATTCTTACACTTCTTTGGGTCCGGGTTAATCCGTTTGTGGCGAAAGGCGGTCCTAT 238

Qy       3217 TCTCGAGATCTGTGGTTTAGACTGCTTGTGATTCGATTGACCGGTGGATGGGTTGGTGAA 3276
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        237 TCTCGAGATCTGTGGTTTAGACTGCTTGTGATTCGATTGACCGGTGGATGGGTTGGTGAA 178

Qy       3277 AAAGGTTTAATTCCCACGGATCAAAGAGAGGTAAGAGAGATATTGTTTTACCTCTAAAAG 3336
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        177 AAAGGTTTAATTCCCACGGATCAAAGAGAGGTAAGAGAGATATTGTTTTACCTCTAAAAG 118

Qy       3337 ACTCCTTCATTGTGTTCATTAGATGATGAAAAATGAAAAGAAAAAGAAGATTTAATTTTG 3396
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        117 ACTCCTTCATTGTGTTCATTAGATGATGAAAAATGAAAAGAAAAAGAAGATTTAATTTTG 58

Qy       3397 TTACGAGAATTGTTATTTTTGCAAGAATGTGTTGTAGATA 3436
             |||||||||||||||||||||||||||||||||||||||
Db         57 TTACGAGAATTGTTATTTTTGCAAGAATGTGTTGTAGATA 18
```

```
RESULT 10
AV564362/c
LOCUS       AV564362                542 bp    mRNA    linear   EST 07-SEP-2000
DEFINITION  AV564362 Arabidopsis thaliana green siliques Columbia Arabidopsis
            thaliana cDNA clone SQ203c01F 3', mRNA sequence.
ACCESSION   AV564362
VERSION     AV564362.1  GI:8735788
KEYWORDS    EST.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
```

```
REFERENCE   1  (bases 1 to 542)
  AUTHORS   Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
  TITLE     A large scale analysis of cDNA in Arabidopsis thaliana: Generation
            of 12,028 non-redundant expressed sequence tags from normalized and
            size-selected cDNA libraries
  JOURNAL   DNA Res. 7, 175-180 (2000)
  MEDLINE   20363093
   PUBMED   10907847
COMMENT     Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .542
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /clone="SQ203c01F"
                     /tissue_type="green siliques"
                     /clone_lib="Arabidopsis thaliana green siliques Columbia"
                     /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT      199 a    116 c     76 g    151 t
ORIGIN

  Query Match             9.3%;  Score 322;  DB 9;  Length 542;
  Best Local Similarity   99.7%;  Pred. No. 2.8e-154;
  Matches  372;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3064 GGTCATCATTCATCTTTACCCGTTCCTTAAAGGTTTGCTTGGGAAACAAGATAGAATGCC 3123
             ||||||||||||||||||||||:||||||||||||||||||||||||||||||||||||||
Db        542 GGTCATCATTCATCTTTACCCGTTCCTTAAAGGTTTGCTTGGGAAACAAGATAGAATGCC 483

Qy       3124 AACCATTATTGTCGTCTGGTCCATCCTCCTGGCCTCGATTCTTACACTTCTTTGGGTCCG 3183
             |||||||||||||||||||||:||||||||||||||||||||||||||||||||||||||
Db        482 AACCATTATTGTCGTCTGGTCCATCCTCCTGGCCTCGATTCTTACACTTCTTTGGGTCCG 423

Qy       3184 GGTTAATCCGTTTGTGGCGAAAGGCGGTCCTATTCTCGAGATCTGTGGTTTAGACTGCTT 3243
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        422 GGTTAATCCGTTTGTGGCGAAAGGCGGTCCTATTCTCGAGATCTGTGGTTTAGACTGCTT 363

Qy       3244 GTGATTCGATTGACCGGTGGATGGGTTGGTGAAAAAGGTTTAATTCCCACGGATCAAAGA 3303
             ||||||||||||||||||||||:||||||||||||||||||:||||||| ||||||||||
Db        362 GTGATTCGATTGACCGGTGGATGGGTTGGTGAAAAAGGTTTAATTCCCCCGGATCAAAGA 303

Qy       3304 GAGGTAAGAGAGATATTGTTTTA:CTTAAAAGACTCCTTTATTGTGTTCATTAGATGAT 3363
             |||||||||||||||||||||||:|||||||||||||||||||||||||||||||||||||
Db        302 GAGGTAAGAGAGATATTGTTTTA:CTTAAAAGACTCCTTTATTGTGTTCATTAGATGAT 243

Qy       3364 GAAAAATGAAAAGAAAAAGAAGATTTAATTTTGTTACGAGAATTGTTATTTTTGCAAGAA 3423
             ||||||||||||||||||||||:||||||||||||||||||:||||||||||||||||||||
Db        242 GAAAAATGAAAAGAAAAAGAAGATTTAATTTTGTTACGAGAATTGTTATTTTTGCAAGAA 183

Qy       3424 TGTGTTGTAGATA 3436
             |||||||||||||
Db        182 TGTGTTGTAGATA 170


RESULT 11
H36985
LOCUS       H36985                   540 bp    mRNA    linear   EST 30-DEC-1997
DEFINITION  15114 Lambda-PRL2 Arabidopsis thaliana cDNA clone 181G15T7, mRNA
            sequence.
ACCESSION   H36985
VERSION     H36985.1  GI:906484
KEYWORDS    EST.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE   1  (bases 1 to 540)
  AUTHORS   Newman,T., deBruijn,F.J., Green,P., Keegstra,K., Kende,H., McIntosh
            ,L., Ohlrogge,J., Raikhel,N., Somerville,S., Thomashow,M., Retzel
            ,E. and Somerville,C.
  TITLE     Genes galore: a summary of methods for accessing results from
            large-scale partial sequencing of anonymous Arabidopsis cDNA clones
  JOURNAL   Plant Physiol. 106, 1241-1255 (1994)
  MEDLINE   95148729
   PUBMED   7846151
COMMENT     Contact: Thomas Newman
            MSU-DOE Plant Research Laboratory
            Michigan State University
            MSU-DOE-PRL, Michigan State University,Plant Biology Bldg.,E.
            Lansing,Mi
            Tel: 517-353-0854
            Fax: 517-353-9168
            Email: 22313tcn@ibm.cl.msu.edu
            Seq primer: T7 dye primer.
FEATURES             Location/Qualifiers
     source          1. .540
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /strain="var columbia"
                     /db_xref="taxon:3702"
                     /clone="181G15T7"
                     /clone_lib="Lambda-PRL2"
                     /note="Vector: lambda Zip-Lox; Site_1: Sal; Site_2: Not;
                     Lambda PRL2 is a cDNA library derived from equal
                     quantities of 4 pools of mRNA.  The mRNA sources were 1) 7
                     day germinated etiolated seedlings; 2) tissue culture
                     grown roots; 3) staged plants half with 24 hour light
                     cycle, half on 16 hr light, 8 hour dark- rosettes;  4)
                     same plants as 3 but aerial tissue (stems, flowers and
                     siliques.  The vector is BRL's lambda Zip-Lox.  The cDNA
                     inserts were directionally cloned with Sal-Not arms using
                     oligo dT primed cDNA.  "
BASE COUNT      106 a    109 c    132 g    172 t     21 others
ORIGIN

  Query Match             9.1%;  Score 313;  DB 14;  Length 540;
  Best Local Similarity   99.7%;  Pred. No. 1.2e-149;
  Matches  363;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2389 TGTTTATTGTACACCAAAGTTAGCGGCTTTCAAAGGATCAGCTCCAATCAATCTTTCGGA 2448
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          28 TGTTTATTGTACACCAAAGTTAGCGGCTTTCAAAGGATCAGCTCCAATCAATCTTTCGGA 87

Qy       2449 TCGTCTCCATCAAGTTCTTCGATGGGCGCTTGGGTCGGTTGAGATTTTCTTGAGTAGGCA 2508
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          88 TCGTCTCCATCAAGTTCTTCGATGGGCGCTTGGGTCGGTTGAGATTTTCTTGAGTAGGCA 147

Qy       2509 TTGTCCTATTTGGTATGGTTATGGAGGTGGGTTGAAATGGCTTGAGCGGTTGTCCTACAT 2568
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         148 TTGTCCTATTTGGTATGGTTATGGAGGTGGGTTGAAATGGCTTGAGCGGTTGTCCTACAT 207

Qy       2569 TAACTCTGTGGTTTACCCGTGGACCTCTCTACCGCTCATCGTTTACTGTTCTCTCCCTGC 2628
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         208 TAACTCTGTGGTTTACCCGTGGACCTCTCTACCGCTCATCGTTTACTGTTCTCTCCCTGC 267

Qy       2629 CATCTGTCTTCTCACTGGAAAATTCATCGTTCCCGAGATTAGCAACTATGCGAGTATCCT 2688
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         268 CATCTGTCTTCTCACTGGAAAATTCATCGTTCCCGAGATTAGCAACTATGCGAGTATCCT 327

Qy       2689 CTTCATGGCGCTCTTCTCGTCGATTGCAATAACGGGTATTCTCGAGATGCAATGGGGCAA 2748
             |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db         328 CTTCATGGCGNTCTTCTCGTCGATTGCAATAACGGGTATTCTCGAGATGCAATGGGGCAA 387

Qy       2749 AGTT 2752
             ||||
Db         388 AGTT 391


RESULT 12
```

```
/529949
OCUS       AV529949                 303 bp    mRNA    linear   EST 01-SEP-2000
EFINITION  AV529949 Arabidopsis thaliana aboveground organs two to six-week
           old Arabidopsis thaliana cDNA clone APZL53c11R 5', mRNA sequence.
CCESSION   AV529949
ERSION     AV529949.1  GI:8690232
EYWORDS    EST.
OURCE      Arabidopsis thaliana (thale cress)
 ORGANISM  Arabidopsis thaliana
           Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
           Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
           ; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
EFERENCE   1  (bases 1 to 303)
 AUTHORS   Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
 TITLE     A large scale analysis of cDNA in Arabidopsis thaliana: Generation
           of 12,028 non-redundant expressed sequence tags from normalized and
           size-selected cDNA libraries
 JOURNAL   DNA Res. 7, 175-180 (2000)
 MEDLINE   20363093
  PUBMED   10907847
OMMENT     Contact: Erika Asamizu
           The First Laboratory for Plant Gene Research
           Kazusa DNA Research Institute
           Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
           Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
EATURES             Location/Qualifiers
     source         1. .303
                    /organism="Arabidopsis thaliana"
                    /mol_type="mRNA"
                    /strain="Columbia"
                    /db_xref="taxon:3702"
                    /clone="APZL53c11R"
                    /tissue_type="aboveground organs"
                    /dev_stage="two to six-week old"
                    /clone_lib="Arabidopsis thaliana aboveground organs two to
                    six-week old"
                    /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                    XhoI"
ASE COUNT       76 a     59 c     67 g    101 t
RIGIN

 Query Match             8.8%;  Score 303;  DB 9;  Length 303;
 Best Local Similarity   100.0%;  Pred. No. 1.4e-144;
 Matches  303;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

y        791 ATCAAATCGAGCAAGATAAATCCTTACCGGATGTTAATTGTGCTACGTCTTGTGATTCTT 850
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
b          1 ATCAAATCGAGCAAGATAAATCCTTACCGGATGTTAATTGTGCTACGTCTTGTGATTCTT 60

y        851 GGTCTCTTCTTTCACTACCGTATTCTTCACCCCGTCAAAGATGCATATGCTTTGTGGCTT 910
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
b         61 GGTCTCTTCTTTCACTACCGTATTCTTCACCCCGTCAAAGATGCATATGCTTTGTGGCTT 120

y        911 ATTTCTGTTATATGTGAGATATGGTTTGCTGTTTCATGGGTTCTTGATCAGTTCCCTAAA 970
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
b        121 ATTTCTGTTATATGTGAGATATGGTTTGCTGTTTCATGGGTTCTTGATCAGTTCCCTAAA 180

y        971 TGGTACCCTATCGAGCGAGAAACGTACTTGGACCGACTCTCATTAAGATATGAGAAAGAA 1030
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
b        181 TGGTACCCTATCGAGCGAGAAACGTACTTGGACCGACTCTCATTAAGATATGAGAAAGAA 240

y       1031 GGGAAACCGTCGGGACTATCCCCTGTGGATGTATTTGTTAGTACAGTGGATCCATTGAAA 1090
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
b        241 GGGAAACCGTCGGGACTATCCCCTGTGGATGTATTTGTTAGTACAGTGGATCCATTGAAA 300

y       1091 GAG 1093
            |||
b        301 GAG 303

ESULT 13
96884
OCUS       N96884                   503 bp    mRNA    linear   EST 05-JAN-1998
DEFINITION  21830 Lambda-PRL1 Arabidopsis thaliana cDNA clone G6E2T7, mRNA
           sequence.
ACCESSION   N96884
VERSION     N96884.1  GI:2748573
KEYWORDS    EST.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 503)
  AUTHORS   Newman,T., deBruijn,F.J., Green,P., Keegstra,K., Kende,H., McIntosh
            ,L., Ohlrogge,J., Raikhel,N., Somerville,S., Thomashow,M., Retzel
            ,E. and Somerville,C.
  TITLE     Genes galore: a summary of methods for accessing results from
            large-scale partial sequencing of anonymous Arabidopsis cDNA clones
  JOURNAL   Plant Physiol. 106, 1241-1255 (1994)
  MEDLINE   95148729
   PUBMED   7846151
COMMENT     On Jan 5, 1998 this sequence version replaced gi:1269003.
            Contact: Thomas Newman
            MSU-DOE Plant Research Laboratory
            Michigan State University
            MSU-DOE-PRL, Michigan State University,Plant Biology Bldg.,E
            Lansing,MI
            Tel: 517-353 0854
            Fax: 517-353-9168
            Email: 22313tcn@ibm.cl.msu.edu
            Seq primer: T7.
FEATURES             Location/Qualifiers
     source          1. .503
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /strain="var columbia"
                     /db_xref="taxon:3702"
                     /clone="G6E2T7"
                     /clone_lib="Lambda-PRL1"
                     /note="Vector: Lambda Shlox-1; Site_1: EcoRI; Site_2:
                     HindIII; Lambda PRL1 is a cDNA library derived from equal
                     quantities of 4 pools of mRNA.  The mRNA sources were 1) 7
                     day germinated etiolated seedlings; 2) tissue culture
                     grown roots; 3) staged plants half with 24 hour light
                     cycle, half on 16 hr light, 8 hour dark- rosettes;  4)
                     same plants as 3 but aerial tissue (stems, flowers and
                     siliques.  The library was made in Novagen's Lambda
                     Shlox-1 with (oligo dT primed) directional inserts cloned
                     between the EcoRI and HindIII sites. "
BASE COUNT      130 a     89 c     130 g     134 t     20 others
ORIGIN

  Query Match             8.6%;  Score 296;  DB 14;  Length 503;
  Best Local Similarity   100.0%;  Pred. No. 7.2e-141;
  Matches  296;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1386 TAAAGATCAATGCTTTAGTAGCAACAGCACAGAAAGTGCCTGAGGATGGTTGGACTATGC 1445
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1 TAAAGATCAATGCTTTAGTAGCAACAGCACAGAAAGTGCCTGAGGATGGTTGGACTATGC 60

Qy      1446 AAGACGGTACACCTTGGCCCGGTAATAGTGTGCGAGATCATCCTGGCATGATTCAGGTCT 1505
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        61 AAGACGGTACACCTTGGCCCGGTAATAGTGTGCGAGATCATCCTGGCATGATTCAGGTCT 120

Qy      1506 TCCTTGGAAGTGACGGTGTTCGTGATGTCGAAAACAACGAGTTGCCTCGATTAGTTTACG 1565
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       121 TCCTTGGAAGTGACGGTGTTCGTGATGTCGAAAACAACGAGTTGCCTCGATTAGTTTACG 180

Qy      1566 TTTCTCGTGAGAAGAGACCCGGATTTGATCACCATAAGAAGGCTGGAGCTATGAATTCCC 1625
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       181 TTTCTCGTGAGAAGAGACCCGGATTTGATCACCATAAGAAGGCTGGAGCTATGAATTCCC 240

Qy      1626 TGATACGAGTCTCTGGGGTTCTATCAAATGCTCCTTACCTTCTGAATGTCGATTGT 1681
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 TGATACGAGTCTCTGGGGTTCTATCAAATGCTCCTTACCTTCTGAATGTCGATTGT 296
```

```
RESULT 14
AV547751/c
LOCUS       AV547751                 263 bp    mRNA    linear   EST 06-SEP-2000
DEFINITION  AV547751 Arabidopsis thaliana roots Columbia Arabidopsis thaliana
            cDNA clone RZL39a02F 3', mRNA sequence.
ACCESSION   AV547751
VERSION     AV547751.1  GI:8719165
KEYWORDS    EST.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 263)
  AUTHORS   Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
  TITLE     A large scale analysis of cDNA in Arabidopsis thaliana: Generation
            of 12,028 non-redundant expressed sequence tags from normalized and
            size-selected cDNA libraries
  JOURNAL   DNA Res. 7, 175-180 (2000)
  MEDLINE   20363093
   PUBMED   10907847
COMMENT     Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .263
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /clone="RZL39a02F"
                     /tissue_type="roots"
                     /clone_lib="Arabidopsis thaliana roots Columbia"
                     /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT       83 a     68 c     28 g     84 t
ORIGIN

  Query Match             7.1%;  Score 245;  DB 9;  Length 263;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-114;
  Matches  245;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3192 CGTTTGTGGCGAAAGGCGGTCCTATTCTCGAGATCTGTGGTTTAGACTGCTTGTGATTCG 3251
             |||| ||||||||||||| ||||||| |||| |||||| ||, ||||| |||| ||||
Db        263 CGTTTGTGGCGAAAGGCGGTCCTATTCTCGAGATCTGTGGTTTAGACTGCTTGTGATTCG 204

Qy       3252 ATTGACCGGTGGATGGGTTGGTGAAAAAGGTTTAATTCCCACGGATCAAAGAGAGGTAAG 3311
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        203 ATTGACCGGTGGATGGGTTGGTGAAAAAGGTTTAATTCCCACGGATCAAAGAGAGGTAAG 144

Qy       3312 AGAGATATTGTTTTACCTCTAAAAGACTCCTTCATTGTGTTCATTAGATGATGAAAAATG 3371
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        143 AGAGATATTGTTTTACCTCTAAAAGACTCCTTCATTGTGTTCATTAGATGATGAAAAATG 84

Qy       3372 AAAAGAAAAAGAAGATTTAATTTTGTTACGAGAATTGTTATTTTTGCAAGAATGTGTTGT 3431
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          83 AAAAGAAAAAGAAGATTTAATTTTGTTACGAGAATTGTTATTTTTGCAAGAATGTGTTGT 24

Qy       3432 AGATA 3436
             |||||
Db          23 AGATA 19


RESULT 15
AV546593/c
LOCUS       AV546593                 380 bp    mRNA    linear   EST 06-SEP-2000
DEFINITION  AV546593 Arabidopsis thaliana roots Columbia Arabidopsis thaliana
            cDNA clone RZL16d10F 3', mRNA sequence
ACCESSION   AV546593
VERSION     AV546593.1  GI:8718007
KEYWORDS    EST.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 380)
  AUTHORS   Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
  TITLE     A large scale analysis of cDNA in Arabidopsis thaliana: Generation
            of 12,028 non-redundant expressed sequence tags from normalized and
            size-selected cDNA libraries
  JOURNAL   DNA Res. 7, 175-180 (2000)
  MEDLINE   20363093
   PUBMED   10907847
COMMENT     Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .380
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /clone="RZL16d10F"
                     /tissue_type="roots"
                     /clone_lib="Arabidopsis thaliana roots Columbia"
                     /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT      134 a     91 c     39 g    116 t
ORIGIN

  Query Match             6.9%;  Score 238;  DB 9;  Length 380;
  Best Local Similarity   100.0%;  Pred. No. 5.8e-111;
  Matches  238;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3199 GGCGAAAGGCGGTCCTATTCTCGAGATCTGTGGTTTAGACTGCTTGTGATTCGATTGACC 3258
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        363 GGCGAAAGGCGGTCCTATTCTCGAGATCTGTGGTTTAGACTGCTTGTGATTCGATTGACC 304

Qy       3259 GGTGGATGGGTTGGTGAAAAAGGTTTAATTCCCACGGATCAAAGAGAGGTAAGAGAGATA 3318
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        303 GGTGGATGGGTTGGTGAAAAAGGTTTAATTCCCACGGATCAAAGAGAGGTAAGAGAGATA 244

Qy       3319 TTGTTTTACCTCTAAAAGACTCCTTCATTGTGTTCATTAGATGATGAAAAATGAAAAGAA 3378
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        243 TTGTTTTACCTCTAAAAGACTCCTTCATTGTGTTCATTAGATGATGAAAAATGAAAAGAA 184

Qy       3379 AAAGAAGATTTAATTTTGTTACGAGAATTGTTATTTTTGCAAGAATGTGTTGTAGATA 3436
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        183 AAAGAAGATTTAATTTTGTTACGAGAATTGTTATTTTTGCAAGAATGTGTTGTAGATA 126


Search completed: October 31, 2003, 09:25:44
Job time : 6766.55 secs
```